VOICE SIGNAL ANALYSIS
USING THE BISPECTRUM

THESIS

Deborah Ann Douglass
Major, Canadian Armed Forces

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GSO/ENG/94D-01

VOICE SIGNAL ANALYSIS USING THE BISPECTRUM

THESIS

Deborah Ann Douglass
Major, Canadian Armed Forces

AFIT/GSO/ENS/94D-01

Approved for public release: distribution unlimited

AFIT/GSO/ENG/94D-01

VOICE SIGNAL ANALYSIS
USING THE BISPECTRUM

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science (Space Operations)

Deborah A. Douglass, B.S.(E.E.)

Major, Canadian Armed Forces

December, 1994

Approved for public release; dsitribution unlimited

Preface

In the initial contemplation of this topic, there seemed to be many papers professing the possible advantages the bispectrum would have in the processing of speech signals. At the time, I did not question the fact there did not seem to be many follow-on papers with clearly defined results. So I eagerly dug in. Thus began many months of frustration and effort. This paper was written to quantify some of the findings that I have come across, and perhaps to lead the next generation on to greener pastures, without falling into the loop holes I found. I would like to take this opportunity to thank, and perhaps appologize to, those who have observed and consoled me in my anguish. Without the friends I have found at AFIT, I might not have made it through with a smile. Best of luck to them in their future posts and may we meet again under more relaxed times.

D.A. Douglass

Table of Contents

List of Figures

Figure

List of Tables

AFIT/GSO/ENG/94D-001

Abstract

The theory of the bispectrum has been studied, though very few practical applications have yet been considered in any depth. One application mentioned in the literature is the use of the bispectrum for voice signal processing. The aim of this thesis was to research the bispectrum towards the particular application of speech enhancement. The technique is based on the fact that the bispectrum is zero for a Gaussian white noise signal, and the bispectrum of two signals added together is the sum of the two signal bispectra. Theoretically, processing signals in the bispectra domain should increase the signal-to-noise ratio of the seech signal. The signal can then be reconstructed from the bispectrum.

Though the theory of the estimation techniques were proven, the applicability of the bispectrum to voice processing was questionable. Since any additive white noise is a random process, it will only be the expected value that is zero. With speech signals, the signal is considered stationary for only approximately 20 milliseconds. This does not allow a significant amont of the noise energy to be removed through the averaging process. Classical methods are just as effective.

VOICE SIGNAL ANALYSIS USING THE BISPECTRUM

I. Introduction

1.1 Background

Communication systems exhibit a number of different kinds of interference, especially when transmitted across vast distances in space. One type of interference is the addition of wideband noise which obscures the original signal. Thus, one problem associated with enhancing speech signals is the attenuation of this wideband random noise, without compromising the intelligibility of the resulting processed speech. Because broadband continuous noise overlaps the speech signal in both time and frequency, it is the most difficult type of noise to remove.

Speech enhancement attempts to improve the performance of voice communications when the signal is corrupted by noise. Systems which operate on the signal only after it has been contaminated by noise primarily improve the quality of the noisy signal at the expense of some intelligibility loss. The quality of a signal is a subjective measure which reflects the way the signal is perceived by listeners. It can be expressed in terms of how pleasant the signal sounds or how much effort is required on behalf of the listeners in order to understand the message. Intelligibility, on the other hand, is an objective measure of the amount of information which can be extracted by listeners from a given signal, whether the signal is clean or noisy. No mathematical quantification of these measures is known.

Traditionally, work on enhancing communications signals has been carried out using second-order statistics, referred to as autocorrelations. The associated Fourier transforms are known as power spectra. The power spectrum, however, does not retain the phase information. Thus, the signal cannot be reconstructed from information obtained through second order statistics. In recent years, several researchers have reported on the use of the third-order statistics, which retain the phase information, for enhancing signals in the presence of additive white noise.

1.2 Statement of Problem

This thesis first investigates the behavior of the bispectrum in the presence of wideband noise and then will consider the use of the bispectrum to remove Gaussian white noise from a signal for speech enhancement.

1.3 Scope

The speech enhancement problem consists of a family of subproblems characterized by the type of noise source, the way the noise interacts with the clean signal, the number of voice channels and the nature of the communication system. In order to limit the scope of this research, the following conditions will be set.

a. Only additive white noise uncorrelated with the speech signal will be considered. This noise is to be continuous and wideband. It is assumed that any impulse or tonal noise can be removed through gating and notch filters.

b. No reference signal to the noise is available, and the clean speech cannot be preprocessed prior to being affected by the noise.

c. Other voice distortions such as speaker stress and fatigue, recording channel distortion, and multipath echo will not be considered.

d. This research will not address the effects due to cochannel interference.

1.4 Summary of Historical Enhancement Approaches

Traditionally, second-order statistics were used to improve the quality of a noisy signal. Two of the major methods are summarized below.

1.4.1 Fourier Transform Approaches. This is a spectral subtraction estimation approach which was developed to enhance speech signals degraded by uncorrelated additive noise. In this method, the power spectrum of the clean signal is estimated by subtracting a scaled version of the noise floor obtained when the signal is not being transmitted. Since the phase cannot be recovered, this magnitude difference is combined with the original noisy phase to form the enhanced speech signal. Using this approach, noise has been reduced by 10 to 15 dB [2:312]. The major drawback of this method of suppressing wideband noise is that it results in noticeable annoying residual noise which consists of narrow band signals with time-varying frequencies and amplitude. This residual noise is often referred to as musical noise [3:1530]. The speech in the enhanced signal is reasonably clear, however, this method does require a reference signal to the noise.

1.4.2 Model-Based Approaches. If the joint statistics of the clean signal and the noisy process are known, a model of the speech can be developed. The speech signal is then estimated by determining the values of the parameters of the model from the noisy signal. There are three basic techniques for the estimation of these parameters:

a. *Maximum Likelihood Estimation* (ML) in which the estimated parameter values are those which would most likely result in producing the observed signal.

b. *Maximum A Posteriori Estimation* (MAP) in which the estimated parameter values are those which maximize the conditional probability of the parameters given the observation.

c. *Minimum Mean Square Estimation* (MMSE) which maximizes the expected value of the conditional probability density of the clean parameters given the noisy observations.

An example of the model-based approaches is the method developed by Lim and Oppenheim [3:1530]. They proposed treating the speech signal as a time varying autoregressive (AR) process estimated from the noisy signal. In their method, speech is modelled as a time-varying AR process whose parameters remain constant over relatively short time frames of 10-20 msec. A time-varying AR model is attributed to the speech signal, and both the model and the signal are estimated from the given noisy signal using the MAP estimation approach. Thus, the vocal tract is considered an all-pole filter for the spectrally flat (Gaussian white noise for the unvoiced signals and an impulse function for one pitch period of voiced signals) excitation signal. This has developed into a method of speech coding known as Linear Predictive Coding (LPC).

1.5 Estimation Using Higher Order Statistics.

During recent years, researchers have found that higher order statistics could be used to enhance signals in the presence of noise. Higher order spectra, defined in terms of higher order moments, known as cumulants, would provide more information than that available from second-order statistics. The Fourier transforms of the cumulants are known as polyspectra. This thesis will concentrate on the third-order spectrum, called the bispectrum. The bispectrum of a process can be defined as the two dimensional Fourier transform of its third moment sequence.

1.5.1 Motivations. In general, there are three motivations behind the use of polyspectra in signal processing. These are:

a. to suppress Gaussian noise processes of unknown spectrum characteristics in detection, parameter estimation and classification problems. A characteristic which differentiates the third-order cumulants from second-order correlation is that the cumulants are blind to all kinds of Gaussian processes, whereas correlation is not. In general terms, the bispectrum has the property of being zero for a Gaussian process. When cumulant-based methods are applied to a non-Gaussian signal, such as a speech signal, that is corrupted by additive Gaussian noise (even colored Gaussian noise), the signal to noise ratio should be improved.

b. to reconstruct the phase and magnitude response of signals or systems. The polyspectrum preserves the true phase character of signals, unlike the power spectrum domain which only preserves the magnitude. The current practice is to process the signal using the second-order statistics, but the noisy phase is carried through for final reconstruction of the signal.

c. to detect and characterize the nonlinear properties of mechanisms which generate time series via phase relations of their harmonic components.

1.5.2 Approach. To process the signals using second-order methods, the initial problem is the estimation of the bispectrum of the signal of interest. As with estimating the power spectra of a signal, there are two general method types used in polyspectral analysis; non-parametric and parametric methods. To date, bispectral analysis of voice signals has been carried out using periodogram-based estimates, which is a non-parametric method [4:IV-488]. Non-parametric methods, however, tend to have a high variance and low resolution. In parametric methods, a model is chosen to approximate the underlying process. The parameters of the model are then estimated for real data observations. The spectrum is computed based on this model. These methods are in the classes of moving average (MA), autoregressive (AR), or autoregressive moving average (ARMA) processes.

Once the bispectrum of a signal is estimated using one of the above methods, the second step is to recover the signal, theoretically stripped of the wideband noise. The most popular method is to use a least squares approach for both the magnitude and phase components of the bispectrum [19:1297]. The bispectrum samples are used to recover the Fourier magnitudes and phases of the signal. It can be shown that there is a linear relationship between the natural log of the bispectrum and the log of the Fourier magnitudes. A least squares solution of each of the samples is taken to recover the Fourier magnitudes. A similar approach can be taken for the Fourier phases. An additional method used in this thesis to recover the phase is to recursively calculate the phase. The signal reconstruction is then carried out by combining the recovered Fourier magnitudes and phase and performing an inverse Discrete Fourier Transform to recover the original signal sequence.

1.6 Thesis Organization

The theory of the bispectrum, its estimation, and the reconstruction of a signal from the bispectrum will be discussed in Chapter II, with particular emphasis on voice signals. In Chapter III, the behavior of the bispectral estimates of Gaussian and Non-Gaussian white noise signals will be considered, as well as the application of the bispectrum to noisy AR processes. The practical application of the bispectrum will continue to its use specifically with voice signals in Chapter IV. The trends discovered will be summarized and discussed in Chapter V.

II. Background Theory

2.1 Introduction

As indicated in Section 1.5.2, the bispectral approach to enhance a signal will be to estimate the bispectrum of the signal and then recover the signal, theoretically stripped of some of the wideband noise energy. This section covers the theory behind this approach. The areas covered will be:

- Bispectrum
- Estimating the bispectrum
- Signal reconstruction
- Voice modelling

2.2 Bispectrum

2.2.1 Defining the Bispectrum. In this section, the definitions and properties of moments, cumulants and higher-order spectra of stationary stochastic signals are introduced first. Since voice contains periodic segments, the theory will be extended to cover the case of periodic processes. Correlations and the power spectra are the most extensively used concepts in the applications of stochastic approaches. These concepts involve only second-order moments. The bispectrum is developed from the same background theories, but it utilizes third-order statistics. The theory of third-order statistics is more fully developed in Appendix A. To summarize the results, if $\{x(k)\}$, $k=-\infty,\ldots,\infty$ is a real, discrete, zero mean, stationary process, then the third-order cumulant is defined to be

$$c_3(m,n)=E[x(k)x(k+m)x(k+n)], \qquad (1)$$

where E is the expectation operator. Since the mean is zero, the third-order moments and cumulants are identical. The bispectrum of this process is defined as the two-dimensional discrete Fourier Transform of the third order-cumulant, or

$$B(\omega_1,\omega_2)=\sum_m \sum_n c_3(m,n)e^{-j(\omega_1 m+\omega_2 n)} \qquad (2)$$
$$|\omega_1|, |\omega_2| \le \pi.$$

2.2.2 Properties of the Bispectrum. In general, the bispectrum is complex and is doubly periodic with period 2π, i.e.,

$$B(\omega_1+2k_1\pi, \omega_2+2k_2\pi)=B(\omega_1, \omega_2) \quad k_1,k_2 \text{ integers.} \qquad (3)$$

The bispectrum has twelve redundant regions which were developed from both the redundant properties of the third moment and the fact that, for real processes, $B(-\omega_1,-\omega_2)=B^*(\omega_1,\omega_2)$. If the bispectrum is defined in any one of these twelve regions, it can be determined everywhere. One such region, is defined by the triangular area bounded by $\omega_1=\omega_2$, $\omega_1=0$, $\omega_1+\omega_2=\pi$, as shown in Figure 1. This is the region that will be used when considering the bispectrum throughout this text.

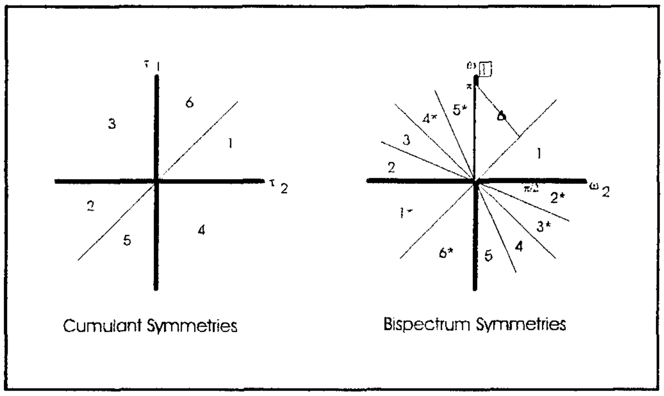


Figure 1 Regions of Symmetry for the third-order moments and the bispectrum. The area of interest is indicated.

Additional properties of the bispectrum that make it very attractive for noisy situations are outlined below. The proofs of these properties are contained in Appendix B.

1) Cumulants are additive, that is the cumulant of two statistically independent random processes that were added equals the sum of the cumulants of the individual random processes. This is true only for third-order moments, it is not true for higher moments.

2) Random processes with symmetric probability density functions have a zero bispectrum. If a random process is symmetrically distributed, then its third-order cumulant equals zero. If we were interested in these signals, the fourth-order cumulants must be used. Examples of symmetrically distributed processes include Laplace, Uniform, Gaussian, and Bernoulli-Gaussian distributions; whereas Exponential, Rayleigh and k-distributions are nonsymmetric.[11:279]. This advantage can be used when dealing with signals with an additive noise signal which has a symmetric probability density function.

3) Non-gaussian white noise processes (with a non-symmetric probability density function) have a flat bispectrum. Just as the covariance function of white noise is an impulse function and its spectrum is flat, the (higher order) cumulants of white noise are multidimensional impulse functions and the polyspectra of this noise are multidimensionally flat. If w(k) is a stationary, non-gaussian process with $E[w(k)]=0$, $E[w(k)w(k+\tau)]=Q\delta(\tau)$, and $E[w(k)w(k+\tau_1)w(k+\tau_2)]=\beta\delta(\tau_1,\tau_2)$, its power spectrum and bispectrum are both flat: $P(\omega)=Q$ and $B(\omega_1,\omega_2)=\beta$.

4) Linear Time Invariant (LTI) systems. The transfer function of the bispectrum for a signal which is passed through an LTI system is:

$$\begin{gathered} B_y(\omega_1,\omega_2)=H(\omega_1)H(\omega_2)H^*(\omega_1+\omega_2)B_x(\omega_1,\omega_2) \\ \text{for } y(k)=\sum_{s=0}^{\infty}h(k-s)x(s) \end{gathered} \tag{4}$$

5) The bispectrum $B(\omega_1,\omega_2)$ and the Fourier Transform $X(\omega_1)$ are related by:

   a. for a periodic signal with a period of N

$$B(\omega_1,\omega_2)=\frac{1}{N}X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2) \tag{5}$$

   b. for an ergodic signal of finite length K

$$B(\omega_1,\omega_2)=\frac{1}{K}X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2)$$

2.2.3 Power Spectrum and the Bispectrum. To show how the power spectrum and the bispectrum are related, each will be looked at in turn. Let $X(\omega)=|X(\omega)|e^{j\theta(\omega)}$ be the Fourier Transform of $x(n)$. If $R(\tau)=E[x(t)x(t+\tau)]$ is the autocorrelation sequence of a ergodic, stationary signal, then the power spectrum is given by

$$\begin{aligned} P(\omega) &= \sum_{\tau}R(\tau)e^{-j\omega\tau} \quad |\omega|<\pi \\ &= X(\omega)X^*(\omega)=|X(\omega)|^2e^{j\alpha(\omega)} \end{aligned} \tag{7}$$

where $\alpha(\omega)=\theta(\omega)+\theta(-\omega)$, and $\theta(\omega)$ is the phase of $X(\omega)$. For a real signal $\theta(\omega)=-\theta(-\omega)$, hence $\alpha(\omega)=0$ [19:705]. Thus it is clear that the Power Spectrum P(ω) does not preserve the phase, which is required for signal reconstruction.

Performing the same construction for the bispectrum, the third moment is given by $R(\tau_1, \tau_2)=E[x(t)x(t+\tau_1)x(t+\tau_2)]$. Using property 5 from the previous section, the bispectrum can be shown to be

$$\begin{aligned} B(\omega_1, \omega_2) &= |B(\omega_1, \omega_2)| e^{j\beta(\omega_1, \omega_2)} \\ &= |X(\omega_1)| \, |X(\omega_2)| \, |X(-\omega_1 - \omega_2)| e^{j(\theta(\omega_1)+\theta(\omega_2)-\theta(\omega_1+\omega_2))}. \end{aligned} \tag{8}$$

From this equation, it is clear that $B(\omega_1,\omega_2)$ preserves the phase information. The information in the phase is important in the processing of speech signals [19:703]. In a later section, methods used to recover the phase and magnitude of the Fourier Transform of x(n) from the bispectrum will be discussed.

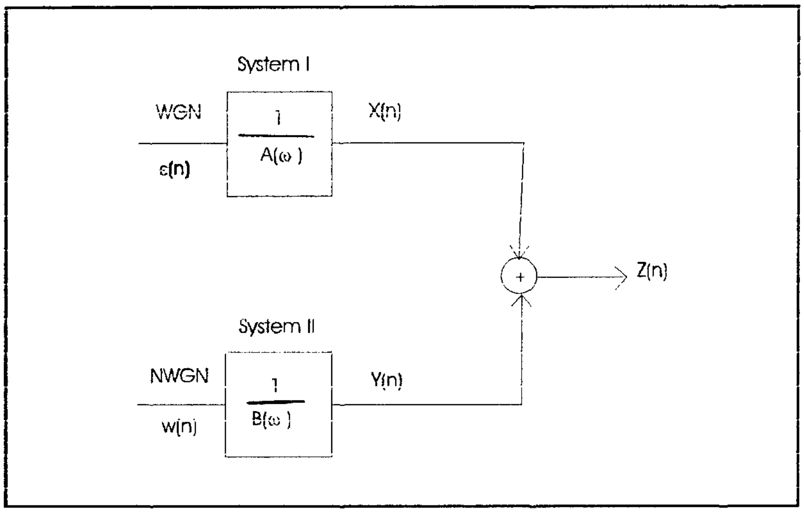


Figure 2 Combining GWN and NGWN fed AR processes. Input processes w(n) and ε(n) are independent.

In some processes, the spectrum and the bispectrum will be modelled by two different linear filters. Consider the process in Figure 2. This process is the sum of two processes, one which is the output of an AR filter driven by GWN and the other is the output of an AR filter driven by NGWN. If the GWN and the NGWN are independent, X(n) and Y(n) are also statistically independent. Thus the bispectrum of Z(n) is the sum of the individual bispectra. Since X(n) is Gaussian, its bispectrum is zero. So the bispectrum of Z(n) is identical to the bispectrum of Y(n), given by

$$B_z(\omega_1,\omega_2)=B_y(\omega_1,\omega_2)=B_w(\omega_1,\omega_2)\frac{1}{A(\omega_1)}\frac{1}{A(\omega_2)}\frac{1}{A^*(\omega_1+\omega_2)}. \tag{9}$$

In contrast, the power spectrum of Z(n) is the sum of the spectrum of X(n) and the spectrum of Y(n), such that

$$\begin{aligned} P_z(\omega) &= P_x(\omega)+P_y(\omega) \\ &= P_\epsilon(\omega)\frac{1}{|A(\omega)|^2}+P_w(\omega)\frac{1}{|B(\omega)|^2} \\ &= \frac{G_\epsilon^2|A(\omega)|^2+G_w^2|B(\omega)|^2}{G_\epsilon^2|A(\omega)|^2*G_w^2|B(\omega)|^2}. \end{aligned} \tag{10}$$

Even if B(ω) is 1 such that the GWN is strictly additive, the power spectrum of Z(n) cannot be used to directly obtain the AR coefficients of A(z). This is the factor that makes obtaining the coefficients in Linear Predictive Coding of voice signals in noisy conditions very inaccurate. Theoretically, the use of the bispectrum to obtain these coefficients will improve the accuracy.

2.3 Estimating the Bispectrum

The general problem of spectral estimation is that of determining the spectral content of a random process based on a finite set of observations from that process. In speech the length of the observation interval is restricted to a limited number of data points by its quasi-stationary nature. A speech sound, or phoneme is stationary for about 20 to 40 milliseconds, depending on the particular phoneme.

There are many recognized ways to estimate the power spectrum, all of which fit into two general classes of estimation techniques: the classical (non-parametric) methods, and the modern (parametric) methods. Although more tedious to calculate than the power spectrum, the bispectrum offers much additional information relating to frequency interaction by coupling. Each of the general methods for the estimation of the bispectrum will be discussed below.

2.3.1 Non-parametric Method. Non-parametric methods of estimating the k^{th} order spectra have long been studied. A periodogram-based method was used by Fulchiero [4] in his paper on the removal of noise from speech signals. The most widely used method is called the direct class [12:348]. This is the method that will be used for comparison purposes in this thesis. In this method, the available set of observations, $\{x(1), x(2),...,x(N)\}$, is segmented into K segments of M samples each. For speech signals, each segment would contain at least one full period for voiced segments. If f_s is the sampling frequency and N_o is the desired number of frequency samples, then $\Delta_o = f_s/N_o$ is the spacing between the frequency samples in the bispectrum domain. The procedure used in directly estimating the bispectrum is as follows:

a. Segment the data into the K segments and subtract the average value of each segment. Each segment is also normalized for comparison purposes by making sure each signal being compared has the same energy level.

b. Assuming that $\{x^i(k),\ k=0,1,2,...,M-1\}$ are the data in segment i, generate the Discrete Fourier Tranform (DFT) coefficients

$$Y^i(\lambda) = \frac{1}{M}\sum_{k=0}^{M-1} x^i(k) e^{-j2\pi k\lambda/M}, \quad \lambda = 0,1,...,\frac{M}{2} \quad i=1,2,...,K. \tag{11}$$

c. Generate the triple product

$$M_3^i(\lambda_1,\lambda_2) = \frac{1}{\Delta_o^2} Y^i(\lambda_1) Y^i(\lambda_2) Y^{i*}(\lambda_1+\lambda_2). \tag{12}$$

d. The moment spectrum of the given data is the average over the K segments

$$M_3(\omega_1,\omega_2) = \frac{1}{K}\sum_{i=1}^{K} M_3^i(\omega_1,\omega_2) \quad \text{where } \omega_i = (2\pi f_s/N_o)\lambda_i \tag{13}$$

Non-parametric polyspectral estimates are subject to the same problems as non-parametric power spectrum methods, that is they are subject to high variance and low resolution. Thus the DFT methods are of limited value for short duration signals because of its relatively low frequency resolution. This is covered in more detail in a later section.

2.3.2 Parametric Models. In parametric modelling, a model is first chosen which would best fit the signal type being estimated. The second step is to estimate the parameters of the underlying data-generating model using a sample of the signal, and then use the model to compute the polyspectrum. There are three main motivations to look for a non-Gaussian, white noise driven parametric model for bispectrum estimation [13:879]:

a) to recover phase information accurately;
b) to increase the resolution capability of an estimator in resolving closely spaced peaks in the bispectrum domain (over the non-parametric methods); and
c) to increase the bispectrum fidelity in the situations where non-Gaussian processes are indeed parametric or may well be approximated by parametric models.

The basis of parametric modeling is the assumption that the signal is predictable from linear combinations of past outputs and inputs. The roots of the numerator and denominator polynomials of the prediction equation make up the zeros and poles of the model respectively. There are two special cases of the model:

a) the all-zero model, known as the Moving Average MA model with: $a_k=0,\ 1<=k<=p$, and
b) the all-pole model, known as the autoregressive (AR) model with: $b_k=0,\ 1<=k<=q$.

As will be discussed in Section 2.5, voice can be modelled using an Autoregressive (AR) model. This process is described by

$$\sum_{i=0}^{p} a_i X(k-i) = W(k), \tag{14}$$

where the W(k) is the input driving sequence. It is important to distinguish between the driving noise of the model, W[n], and any observation noise. As will be seen in Section 2.5, speech is commonly recognized to be the result of a linear filter fed either by an impulse or white noise. From Property 4 of a bispectrum, which demonstrates the transfer function of a signal passed through a LTI system, the output was shown to be

$$B_y(\omega_1,\omega_2) = H(\omega_1)H(\omega_2)H^*(\omega_1+\omega_2)B_x(\omega_1,\omega_2). \tag{15}$$

For the impulse input, it can be shown that

$$\begin{aligned} B_i(\omega_1,\omega_2) &= \sum_m \sum_n E[G\delta(k)G\delta(k+n)G\delta(k+n)]e^{-j(\omega_1 m+\omega_2 n)} \\ &= G^3 \end{aligned} \tag{16}$$

and for the white noise

$$B_{WN}(\omega_1,\omega_2)=\sum_m\sum_n E[Gw(k)Gw(k+m)Gw(k+n)]e^{-j(\omega_1 m+\omega_2 n)}$$
$$=G^3 \tag{17}$$

since $w(k)$, $w(k+m)$, and $w(k+n)$ are uncorrelated except for when $m=n=0$. Thus the two results are identical and the relations linking the third order moment coefficients of the output of an all-pole filter are the same whether the input is a single impulse or white noise.[1] Due to this property, both of these inputs can be modelled using an input sequence made up of independent, identically distributed, random variables with zero mean and

$$E[W(k)W(k+m)W(k+n)]=\beta\delta(m,n). \tag{18}$$

Such an input is considered a third-order white sequence, but it is non-gaussian. For the process given in Equation 14, the triple correlation sequence becomes

$$C(m,n)+\sum_{i=1}^{p}a_i C(i+m,i+n)=\beta\delta(m,n) \qquad m,n\geq 0 \tag{19}$$

where $C(m,n)$ is the third moment of the AR process. This relationship is proven in Appendix C.

To estimate the bispectrum, the third order cumulants of the (possibly noisy) observation data are estimated first and an AR model is then fitted. Cumulants involve expectations, and as in the case of correlations, they cannot be computed in an exact manner from real data; they must be approximated. For a stationary (and ergodic) process, the autocorrelation can be computed as a time average, ie.

$$C_3(m,n)\approx(\frac{1}{N_R})\sum_{t\in R}x(t)x(t+m)x(t+n) \tag{20}$$

where N_R is the number of samples in sequence R [11].

Non-stationary processes are not ergodic; therefore, a time average cannot substitute for the ensemble average. However, for a certain class of non-stationary processes, known as locally

[1]The corresponding result for second-order systems was shown by Makhoul [8:565]. He showed that both types of input have identical autocorrelations and identical flat spectra.

stationary processes, it is reasonable to estimate the autocorrelation function with respect to a point in time as a short-time average. Thus, as for a stationary (and ergodic) process the autocorrelation can be computed as a time average. An example of a quasi-stationary processes that can be considered locally stationary is speech.

Continuing from Equation 15, the higher-order spectra of the system input and output are related by

$$B_{out}(\omega_1,\omega_2)=G^3H(\omega_1)H(\omega_2)H^*(\omega_1+\omega_2), \tag{21}$$

where

$$H(\omega)=\frac{1}{\sum_{i=0}^{p}a_i e^{-j\omega_i}};\quad a_0=1.$$

is the transfer function of the AR system.

Several methods have been suggested for determining the coefficients $\{a_i\}$ of the AR model. The *Third-Order Recursion* (TOR) is considered one of the simplest methods. In this method the triple correlation equations are solved using a biased estimate of the third-order moments. The process data is windowed, which diminishes the bispectral resolution. The method used in this research, which is called the *Constrained Third-Order Mean* (CTOM), eliminates the windowing [16:1215]. This method substitutes estimated third order moments in place of the true moments which are not known, and is carried out as follows:

a) Form the biased third moment estimates.
   1) Segment the data into K records of M samples each.
   2) For each record obtain $c^i(m,n)$, the biased estimated of the third moment as

$$c^i(m,n)=\frac{1}{M}\sum_{l=\max(1,1-m,1-n)}^{\min(M,M-m,M-n)} X^i(l)X^i(l+m)X^i(l+n)\quad i=1,\dots,K. \tag{23}$$

   3) Average $c^i(m,n)$ over all records to obtain the overall estimate $\hat{C}(m,n)$ of $C(m,n)$ as

$$\hat{C}(m,n)=\frac{1}{K}\sum_{i=1}^{K}c^{i}(m,n). \tag{24}$$

b) Substitute the estimated moments in place of the true ones in Equation 19. By using the 2p+1 third moments values of the m=n line, the matrix equation

$$\hat{C}\alpha=\hat{b} \tag{25}$$

can be formed where

$$\hat{C} \triangleq \begin{bmatrix} \hat{C}(0,0) & \hat{C}(1,1) & \dots & \hat{C}(p,p) \\ \hat{C}(-1,-1) & \hat{C}(0,0) & \dots & \hat{C}(p-1,p-1) \\ \vdots & & & \\ \hat{C}(-p,-p) & \hat{C}(-p+1,-p+1) & \dots & \hat{C}(0,0) \end{bmatrix}, \tag{26}$$

$\alpha \triangleq [1,a_1,\dots,a_p]^T$ are the estimates of the AR parameters, and $\hat{b}=[\beta,0,\dots,0]^T$ is the estimate of the third moment of the driving noise. A least squares solution is found for the matrix equation, which is

$$\alpha=(\hat{C}^T\hat{C})^{-1}\hat{C}^T\hat{b}. \tag{27}$$

c) Form the bispectrum estimate by substituting the estimated coefficients in Equation 21.

2.3.3 Statistical Properties of Estimators. It has been shown [13:878] that the conventional estimators are asymptotically unbiased and consistent with asymptotic variances given by

$$var[R_e\hat{B}(\omega_1,\omega_2)]=var[I_m\hat{B}(\omega_1,\omega_2)]\simeq\frac{1}{KM}P(\omega_1)P(\omega_2)P(\omega_1+\omega_2) \tag{28}$$

where P(ω) is the true power spectrum of the process. Conventional estimators are generally of high variance and therefore a large number of records (K) is required to obtain smooth bispectral estimates. In the case of "short" data records, the K could be increased by using overlapping records.

For parametric modelling, the determination of the statistics of the AR parameter estimators is formidable, if not impossible. There are not even results available for the estimation of the second-order power spectrum [5:190]. Thus far, approximations based on large sample theory have been relied on. In another section of this paper, the statistics of the estimation techniques will be considered.

2.3.4 Noisy Conditions. The calculation of the AR estimators using the power spectrum in low SNR is very inaccurate. The reason for this degradation is that the all-pole model assumed in AR spectral estimation is no longer valid when observation noise is present. If y[n] denotes the noise corrupted AR process x[n], and w[n] denotes the observation noise, the observed process is $y[n]=x[n]+w[n]$. Assuming that the noise is white with variance σ_w^2 and is uncorrelated with x[n], the PSD is

$$\begin{aligned} P_{yy}(\omega) &= \frac{\sigma^2}{A(\omega)A^*(\omega)} + \sigma_w^2 \\ &= \frac{\sigma^2 + \sigma_w^2 A(\omega)A^*(\omega)}{A(\omega)A^*(\omega)}. \end{aligned} \tag{29}$$

Thus the PSD of $y[n]$ is characterized by zeros as well as poles and the AR model is not appropriate for the observed process. This inconsistency of the AR model for a noise corrupted AR process leads to the degradation of the AR estimation [5:201].

The bispectrum of GWN, on the other hand, has an expected value of zero. Thus the bispectrum of $y[n]$ is the same as the bispectrum of $x[n]$ and remains an AR process. This should reduce the degradation of the AR estimation in noisy conditions.

2.4 Signal Reconstruction

If the signal of interest, x(n), is real, band-limited and of finite duration, then x(n) can be reconstructed from its bispectrum information except for a time shift. This is equivalent to recovering the Fourier Transform X(ω) except for a linear phase factor. That is, if $x_M(k)=x(k-s)$ where s is a constant integer, then in the frequency spectrum $X_M(\omega)=X(\omega)e^{2\pi j\omega s}$, and

$$\begin{aligned} B_M(\omega_1,\omega_2) &= X_M(\omega_1)X_M(\omega_2)X_M^*(\omega_1+\omega_2) \\ &= X(\omega_1)e^{2\pi js\omega_1}X(\omega_2)e^{2\pi js\omega_2}X^*(\omega_1+\omega_2)e^{2\pi js(-\omega_1-\omega_2)} \\ &= B(\omega_1,\omega_2). \end{aligned} \tag{30}$$

In contrast, work based on second-order statistics such as autocorrelation is completely "phase-blind" in that the phase information is not retained. In many applications, the phase information is important. One such application area is speech processing.

Once the bispectrum of a signal is estimated using one of the methods in Section 2.3, the second step is to recover the signal, theoretically stripped of the wideband noise. The basis of reconstructing a signal from the bispectrum is the recovery of the Fourier Transform phase and magnitude information. One method [20:1297] is to use a least squares approach for both the magnitude and phase components of the bispectrum. It can easily be shown that there is a linear relationship between the natural log of the bispectrum and the log of the Fourier magnitudes. A least squares solution of each of the samples is taken to recover the Fourier magnitudes. A similar approach can be taken for the Fourier phases. The signal reconstruction is performed by combining the recovered Fourier magnitudes and phase and performing an inverse Discrete Fourier Transform (DFT) to recover the original signal sequence.

Assuming that the signal input, $\{W(n)\}$ has non-zero skewness, i.e. $E[W^3(n)]=\beta\neq 0$, then the bispectrum contains the non-minimum phase information. The bispectrum is given by [13:885]

$$B(\omega_1,\omega_2)=|B(\omega_1,\omega_2)|e^{j\psi(\omega_1,\omega_2)} \tag{31}$$

where

$$|B(\omega_1,\omega_2)|=|X(\omega_1)|\;|X(\omega_2)|\;|X(\omega_1+\omega_2)| \tag{32}$$

$$\psi(\omega_1,\omega_2)=\theta(\omega_1)+\theta(\omega_2)-\theta(\omega_1+\omega_2) \tag{33}$$

and

$$X(\omega)=|X(\omega)|e^{j\theta(\omega)}. \tag{34}$$

2.4.1 Magnitude Reconstruction. The magnitude of the DFT of the signal can be reconstructed from the bispectrum using a Least Squares Approach [20:1299]. This method utilizes the majority of the bispectrum samples which lie in the principal region, as defined in Section 2.2.2. Equation 32 forms the basis for magnitude recovery. Consider a signal of length N. By taking the natural logarithm of both sides, the bispectrum can be expressed as a linear combination of the log of the Fourier magnitudes. Thus we can write

$$\tilde{B}(k,l)=\tilde{X}(k)+\tilde{X}(l)+\tilde{X}(k+l) \tag{35}$$

where

$$\tilde{X}(k)=ln(|X(k)|)$$
$$\tilde{B}(k,l)=ln(|B(k,l)|).$$

This linear combination can be expressed in matrix form as

$$\tilde{b}=A\tilde{x} \tag{36}$$

where

$$\tilde{b}=[\tilde{B}(1,1),\tilde{B}(2,1),...,\tilde{B}(N/2-1,1),\tilde{B}(2,2),\tilde{B}(3,2),...,\tilde{B}(N/4,N/4)]^T \tag{37}$$

is a vector of dimension $(N^2/16)x1$;

$$\tilde{x}=[\tilde{X}(1),\tilde{X}(2),...,\tilde{X}(N/2)]^T \tag{38}$$

is a vector with dimension $(N/2)x1$; and

$$A=\begin{pmatrix} 2 & 1 & 0 & 0 & 0 & ... & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & ... & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & ... & 0 & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & 0 & 0 & 0 & 0 & ... & 1 & 1 \\ 0 & 2 & 0 & 1 & 0 & ... & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & ... & 0 & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix} \tag{39}$$

is a matrix of dimensions $(N^2/16)x(N/2)$. The matrix A is full rank for all orders N and is pseudo-invertible; therefore, the solution for $X(\tau)$ can be obtained using the least squares method for overdetermined matrices:

$$\tilde{x}=(A^TA)^{-1}A^T\tilde{b}. \tag{40}$$

Note that the proposed method requires that N≥8 to have an overdetermined condition. The Fourier magnitudes of the sequence can be found by taking the exponential of each X(τ) i.e.,

$$|X(k)|=exp[\tilde{X}(k)]; \quad k=1,2,...,\frac{N}{2} \tag{41}$$

Since the Fourier sequence is conjugate symmetric,

$$|X(k)|=|X(N-k)|;k=\frac{N}{2}+1,...,N-1. \tag{42}$$

2.4.2 Phase Reconstruction. Several methods have been suggested for phase estimation in the bispectrum domain when generating the bispectrum estimates via non-parametric methods. When computing the non-parametric estimates, a non-wrapped phase can be determined. A non-wrapped phase has not been restricted to a value between +π and -π, but is left as the strict addition of the respective Fourier phases, as given in Equation 33. This non-wrapped phase can be used to reconstruct the Fourier phase in a least squares method similar to that used to obtain the magnitude values. In parametric methods, or when averaging is carried out, only wrapped phase values are available, thus the least squares method cannot be used to reconstruct the phase. A method that was used in the practical applications of this thesis to overcome the phase unwrapping problem was to reconstruct the phase recursively using exponential values. Bartlett et al [1] gave a method of recursive reconstruction of the fourier phase from the bispectra. This method was adapted for use when only wrapped phases were available by taking the exponential of the phase values, since the exponential is indifferent to the 2π phase shifts inherent in phase wrapping.

By rearranging Equation 33 and taking the exponential of each side such that:

$$e^{j\theta(\omega_1 \quad \omega_2)} \quad e^{j[\theta(\omega_1) \quad \theta(\omega_2) \quad \psi(\omega_1, \quad _2)]} \tag{43}$$

the phase can be calculated recursively. Several recursive steps for the phase computation are listed in Table 1. The θ_p can be determined for $p>1$. The values for θ_0 and θ_1 must be arbitrarily chosen. Since the signal has zero mean (zero DC level), θ_0 can be set to zero without affecting the results. The value chosen for θ_1 would affect the signal by imposing a linear phase factor which ultimately corresponds to a time shift in the signal.

Table 1 shows that the recovered phase values can have several independent representations. This fact can be applied to improve the signal-to-noise ratio of the reconstruction by averaging the different results in the case of a noisy bispectrum [1:3125]. The order of calculations given in Table 1 has the advantage that all representations of θ_p can first be averaged before the value itself is used to calculate a further value. Note again that exponential factors should be averaged instead of using direct summation because the θ_p are only determined up to modulo 2π.

Table 1 Recursive Phase Recovery from a Bispectrum. This table shows the recursive steps followed to recover the Fourier phases. Note that the exponential is not indicated.

		Phases used	Phase Recovered
p	q	$\theta_p+\theta_q-\psi_{p,q}$	θ_{p+q}
1	1	$\theta_1+\theta_1-\psi_{1,1}$	θ_2
2	1	$\theta_2+\theta_1-\psi_{2,1}$	θ_3
2	2	$\theta_2+\theta_2-\psi_{2,2}$	θ_4
3	1	$\theta_3+\theta_1-\psi_{3,1}$	θ_3
3	2	$\theta_3+\theta_2-\psi_{3,2}$	θ_5
3	3	$\theta_3+\theta_3-\psi_{3,3}$	θ_6
4 ⋮	1 ⋮	$\theta_4+\theta_1-\psi_{4,1}$ ⋮	θ_5 ⋮
n-2	1	$\theta_{n-2}+\theta_1-\psi_{n-2,1}$	θ_{n-1}
n-1	1	$\theta_{n-1}+\theta_1-\psi_{n-1,1}$	θ_n

Once the Fourier phase and magnitude are obtained, reconstruction is performed by combining the recovered Fourier phases and magnitudes and performing an inverse DFT to recover the original signal.

2.5 Voice Modelling

In order to determine how speech can best be modelled, the basic concepts of how the human vocal system works must first be introduced. The vocal organs work by using compressed air supplied by the lungs through the trachea. With some sounds, the compressed air is be subjected to periodic pulses by the vocal cords (the glottis). The repetition rate of these pulses forms the pitch. A signal which follows this process has a periodic form and is termed voiced. Another sound type is formed when the compressed air passing through the vocal cords is not periodically excited. It is, instead, forced to pass through a small opening, causing an air turbulence to occur. This generates a broadband noise-like sound. This speech is termed unvoiced. After passing through the glottal output, the speech sound, voiced or unvoiced, is subject to a filtering operation by the shape of the vocal tract. Thus, the vocal tract acts as an acoustical tube which strongly passes some natural frequencies, called formants.

In the case of voiced sounds, the pitch frequency and the first, second and third formant frequencies are normally located below the 3 kHz frequency. Voiced speech is characterized by a periodic behavior where the fundamental frequency and the pitch frequency may range from 30 Hz to about 500 Hz. This pitch varies between males and females. The normal pitch frequency is around 125 Hz. Unvoiced speech, on the other hand, has virtually no periodicity and behaves like wide-band noise with less energy than voiced speech.

Many systems directed at processing speech rely, at least to some extent, on a model of the speech waveform as a response of the vocal tract, represented as a quasi-stationary linear system. The linear system reacts to an input of a pulse-train excitation for voiced sounds or a noise-like excitation for unvoiced speech. To produce a continuous, speech-like signal, the mode of excitation and the resonance properties of the linear system must change with time. Thus, speech can be represented as the output of a linear, time varying system whose properties vary slowly with time. If we consider short segments of the speech signal, then each segment can effectively be modelled as a linear time-invariant system, excited either by a quasi-periodic impulse train or a random noise signal. Such a model is illustrated in Figure 3.

This method of modeling speech signals is a parametric method. In this method, an appropriate model is chosen and the parameters for that model are estimated. Many discrete-time random processes encountered can be approximated by such a *time series or rational transfer*

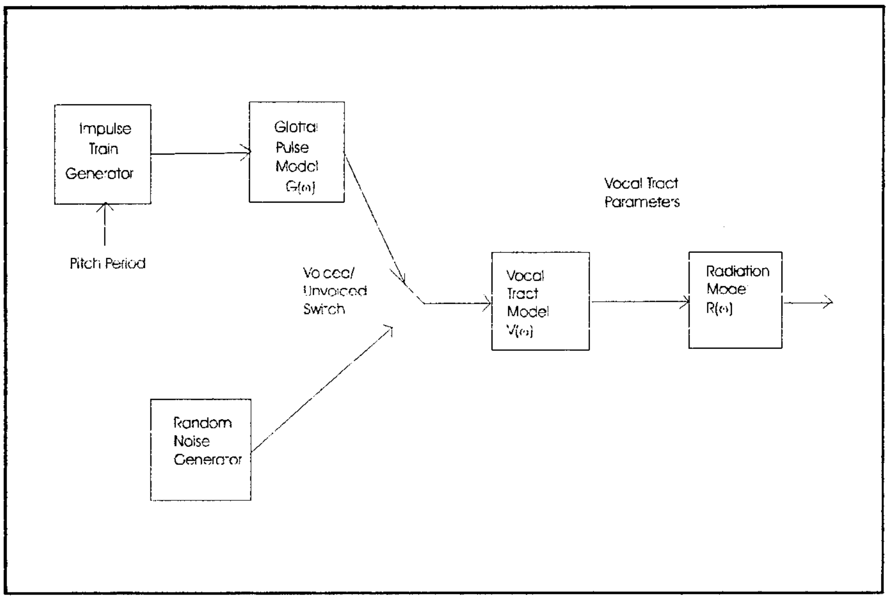


Figure 3 Representative model of the Human Speech Process

function model. In this model, an input driving sequence, $u[n]$, and the output sequence, $x[n]$, that is to model the data are related by the linear difference equation,

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + \sum_{k=0}^{q} b[k]u[n-k]. \tag{44}$$

This general model is termed an Autoregressive, Moving Average (ARMA) model. It must be remembered that the driving noise of the model, $u[n]$, is different from any observation noise. It is, instead, an innate part of the model and gives rise to the random nature of the observed process $x[n]$.

The system transfer function between the input $u[n]$ and the output $x[n]$ of such a system is the rational function [5:109]

$$H(e^{j\omega k}) = \frac{B(e^{j\omega k})}{A(e^{j\omega k})}$$
$$\text{where } A(e^{j\omega k}) = \sum_{k=0}^{p} a[k] e^{-j\omega k}, \quad a_0 = 1, \tag{45}$$
$$B(e^{j\omega k}) = \sum_{k=0}^{q} b[k] e^{-j\omega k}.$$

In the model of a speech signal used in this thesis, only the pole portion, or the Autoregressive (AR) part of the model will be used. Thus, the composite spectrum effects of radiation, vocal tract, and glottal excitation are represented by a time-varying digital filter whose steady-state function is of the form

$$H(e^{j\omega k}) = \frac{G}{1 + \sum_{k=1}^{p} a_k e^{-j\omega k}} \tag{46}$$

To form the model, the resonances (formants) of speech correspond to the poles of the transfer function. An all-pole model is a very good representation of vocal tract effects for a majority of speech sounds; however, acoustic theory indicates that to fully model nasal and fricative sounds, both poles and zeros are required [15]. To accommodate these cases, either zeros must be included in the transfer functions or it may be reasoned that the effect of a zero of the transfer function can be achieved by including more poles. This is the approach to be taken for the speech model in this study. A rule of thumb developed by Parson [19] gives the number of poles p required to model the voice signal as

$$p = \frac{f_s}{1000} + \gamma \tag{47}$$

where f_s is the sampling frequency of the original data and γ is a "fudge factor" (typically 2 or 3) for adding extra poles to the model for flexibility.

In continuous sounds such as vowels, the parameters change very slowly and the model should work very well. With transient stops, the sound, and thus the parameters, change more quickly. The model will theoretically not be as good. The use of transfer functions implicitly assumes that we can represent the speech signal as a stationary signal on a "short-time" basis. That is, the parameters of the model are assumed to be constant over time intervals typically 10-20 msec long [15]. This is termed quasi-stationary.

Thus with a transfer function as given by Equation 46 the speech waveform is assumed to satisfy a difference equation of the form

$$s(n) = \sum_{k=1}^{P} a_k s(n-k) + u(n) + e(n) \qquad (48)$$

on a short time basis where $u(n)$ is the input excitation to the system and $e(n)$ represents the modeling error. For unvoiced speech, $u(n)$ is random noise. For voiced speech, $u(n)$ over each analysis frame consists of one or several impulses with spacing corresponding to the fundamental pitch period. Since u(n) has the specific form of one or several impulses, thus its influence on the estimation procedure is minor. This is substantiated experimentally by virtue of the fact that with the excitation treated as random, one set of estimation procedures corresponds exactly to linear prediction analysis which is well known to be successful for both voiced and unvoiced speech [6:198].

2.6 Summary

This chapter presented the theory of the bispectrum. This background theory was used to show the development of methods used to estimate the bispectrum and to reconstruct a signal from its bispectrum. These methodologies will be used in the next sections in which the bispectrum of different signal types are researched, with special emphasis given to the bispectrum of speech signals.

III. Bispectrum of White Noise Signals

3.1 Introduction

In Chapter II the basic concepts and theory of the bispectrum were discussed. This chapter presents the behavior of the estimated bispectrum, comparing the estimates to expected results. In the first section, the bispectrum of gaussian and non-gaussian white noise is estimated by parametric and non-parametric methods. These two signal types are then used as input into an Autoresgressive (AR) 4 process and the bispectrum of the output is again compared to the theoretical results. Next, the number of averages required to get a good non-parametric estimate is considered. In the last section of this chapter, the estimated bispectrum of a signal consisting of an AR4 process with a non-gaussian input and wideband additive gaussian white noise is researched.

3.2 Bispectrum of Gaussian and Non-Gaussian Signals

To study the use of the bispectrum for the analysis of speech signals, the first consideration was to determine the bispectrum of each of the contributing signals, the white gaussian noise signal and the non-gaussian white noise signal used to feed the speech model. The MATLAB normal distribution random number generator was used to obtain the gaussian white noise (GWN) signal. The non-gaussian white (NGWN) signal was generated by passing a zero-mean white gaussian sequence $\varepsilon(n)$ through a nonlinear filter as follows:

$$W(n)=\begin{cases}\epsilon(n) & \text{if } \epsilon(n)\geq 0\\ 0 & \text{otherwise.}\end{cases} \tag{49}$$

To standardize the signals being compared and to ensure a zero-mean signal, the mean value was subtracted from W(n), which was then divided by its standard deviation. A sample of these signals are shown in Figure 4. To estimate the bispectrum for each signal type, non-parametric and a parametric methods as given in Section 2.3 were carried out on 16 segments of 128 samples. A Hamming window was applied to each segment before estimating to reduce the spectral leakage associated with finite observation intervals.

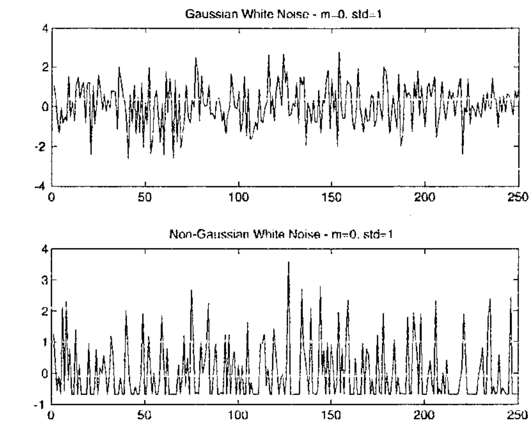


Figure 4 Sample segments of the two signal inputs under consideration: a) Gaussian White Noise, b) Non-Gaussian White Noise

For gaussian white noise, $E[x(k)x(k+\tau_1)x(k+\tau_2)]=0$, therefore the bispectrum is expected to be zero. The estimated bispectra are shown in Figure 5. For the non-gaussian signal, $E[x(k)x(k+\tau_1)x(k+\tau_2)]=G^3\delta(\tau_1,\tau_2)$, so the bispectrum is expected to be constant at G^3. The gain, G^3, is analogous to the energy of a power spectrum and will be called the 3D-energy for the purposes of this thesis. This value can be estimated by the triple correlation estimate $\hat{c}(0,0)$. The 3-D energy of the NGWN signal used was 1.8049 (or a log magnitude value of 0.2565). The non-parametric estimate is affected by the energy loss due to windowing the data. After windowing, the 3-D energy estimate was 0.4977 (or a log magnitude value of -0.3031). The results are graphed in Figure 6.

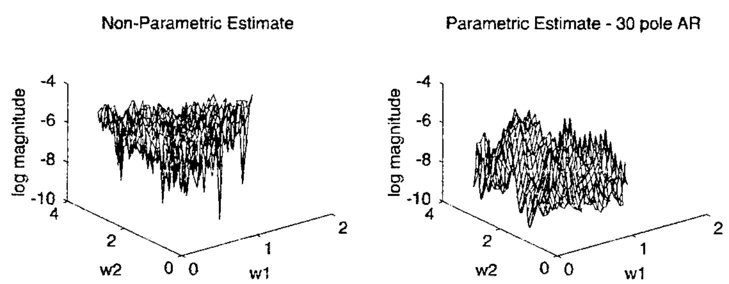


Figure 5 Estimation of the Bisectrum of Gaussian White Noise

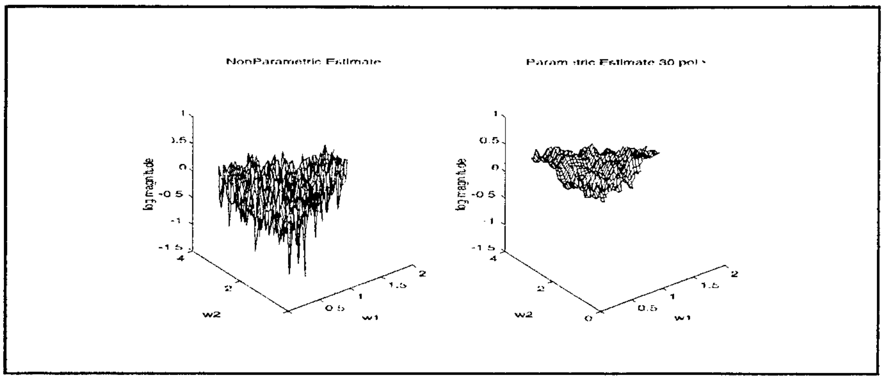


Figure 6 Estimation of the Bispectrum of a Non-Gaussian White Noise Signal

3.3 Fourth-Order Autoregressive Model with Gaussian and Non-Gaussian Input Signals

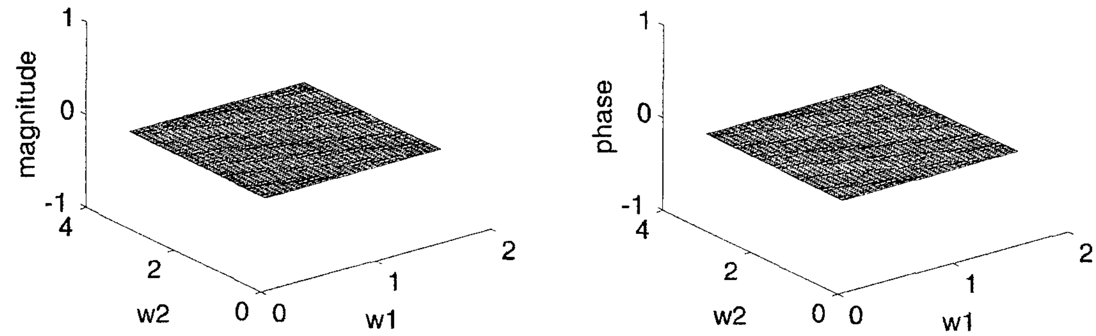
In the second step of the development, a fourth-order AR process with a smooth bispectrum was considered. The process used was

$$X(n) + \sum_{i=1}^{4} a(i)X(n-i) = W(n) \tag{50}$$

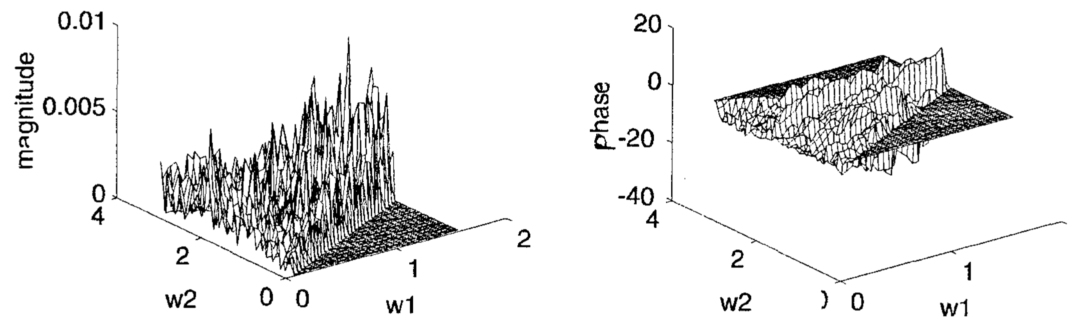
where a(1)=0.1, a(2)=0.2238, a(3)=0.0844, and a(4)=0.0294. The coefficient values were chosen to give a smooth bispectrum. W(n) was either a zero-mean gaussian white noise process (GWN) or a zero mean non-gaussian white process (NGWN) of the form given in Section 3.1. The third order moments were again estimated using 16 records of 128 samples each. Initially a fourth order model was used to obtain the parametric estimation. The expected and estimated results are shown in Figures 7 and 8. The expected bispectrum can be determined using Property 4 using the theoretical bispectrum as the input, which is flat at a value of the 3-D energy for the non-gaussian white noise signal. The expected bispectrum of the NGWN fed process is shown using a unity 3-D energy. The amplitude of the estimates reflect the actual 3-D energy of the signals tested. This 3-D energy can be estimated by

$$\frac{\sum_{i=1}^{N} x^3(i)}{N}, \tag{51}$$

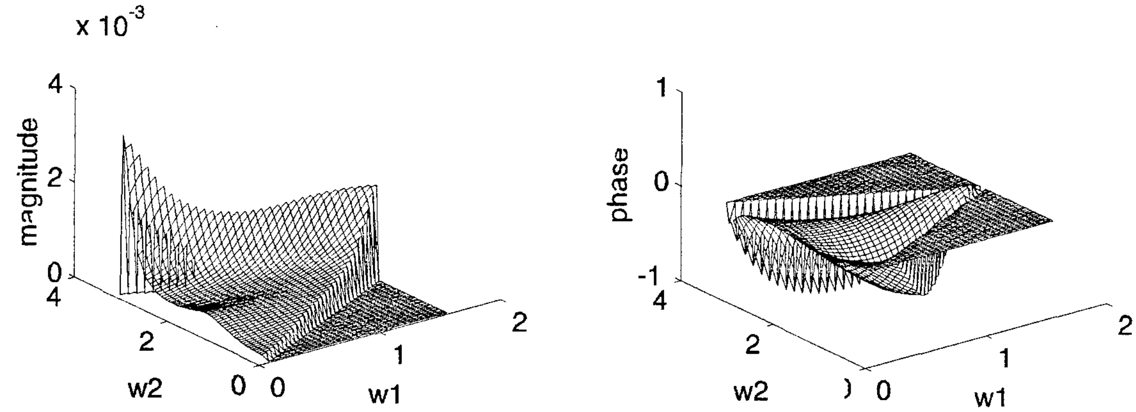
which was 1.55 for the NGWN process and 0.04 for the GWN process. The energy loss due to windowing was compensated for in the case of the non-parametric estimate. Despite this magnitude change, it can be seen that the non-parametric estimates based on 16 representations bear little resemblance to the true bispectrum due to its high variance. To reduce this variance,



a) True Bispectrum

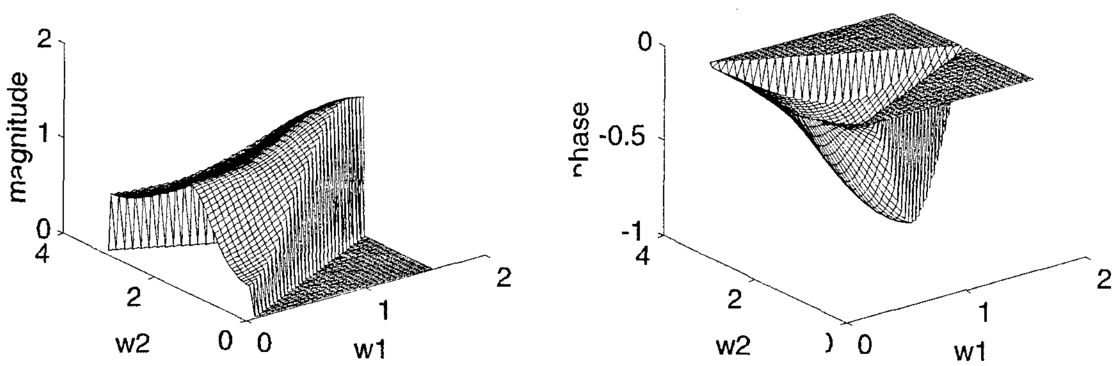


b) Non-Parametric Estimate

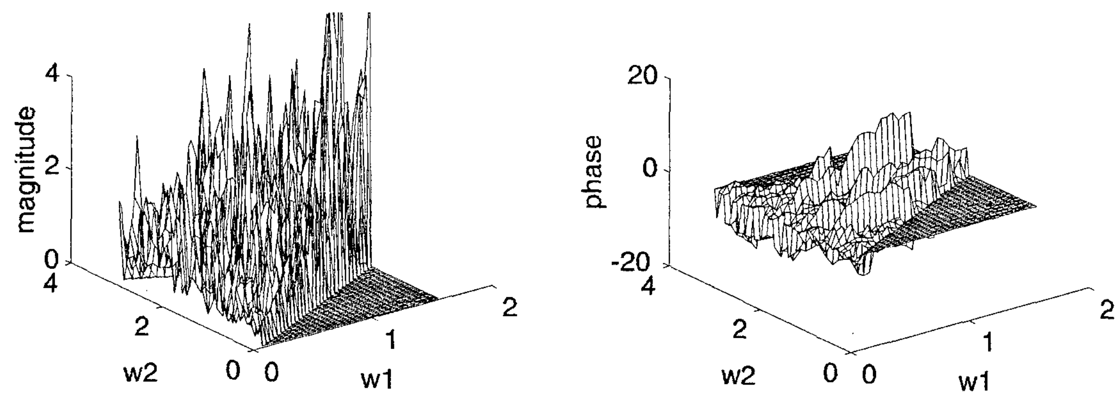


c) Parametric Estimate using 4 Poles

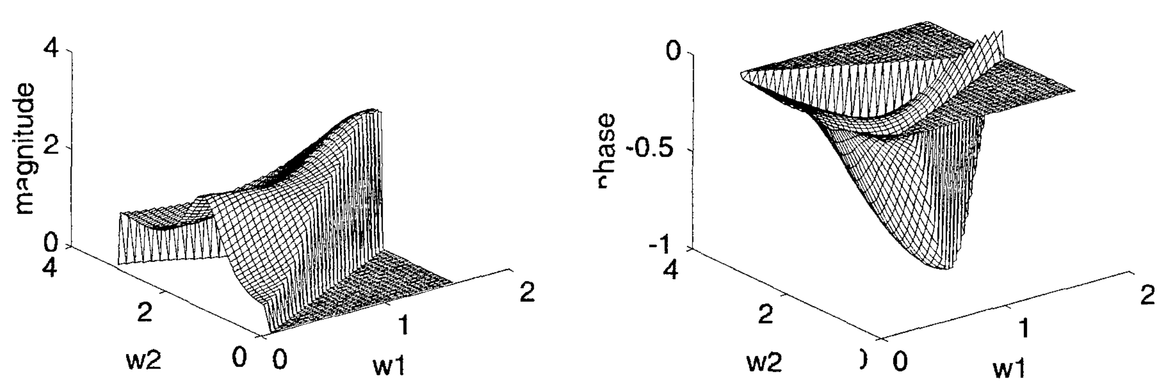
Figure 7 Bispectrum of a Fourth-Order Autoregressive Model with Gaussian White Noise Input.



a) True Bispectrum

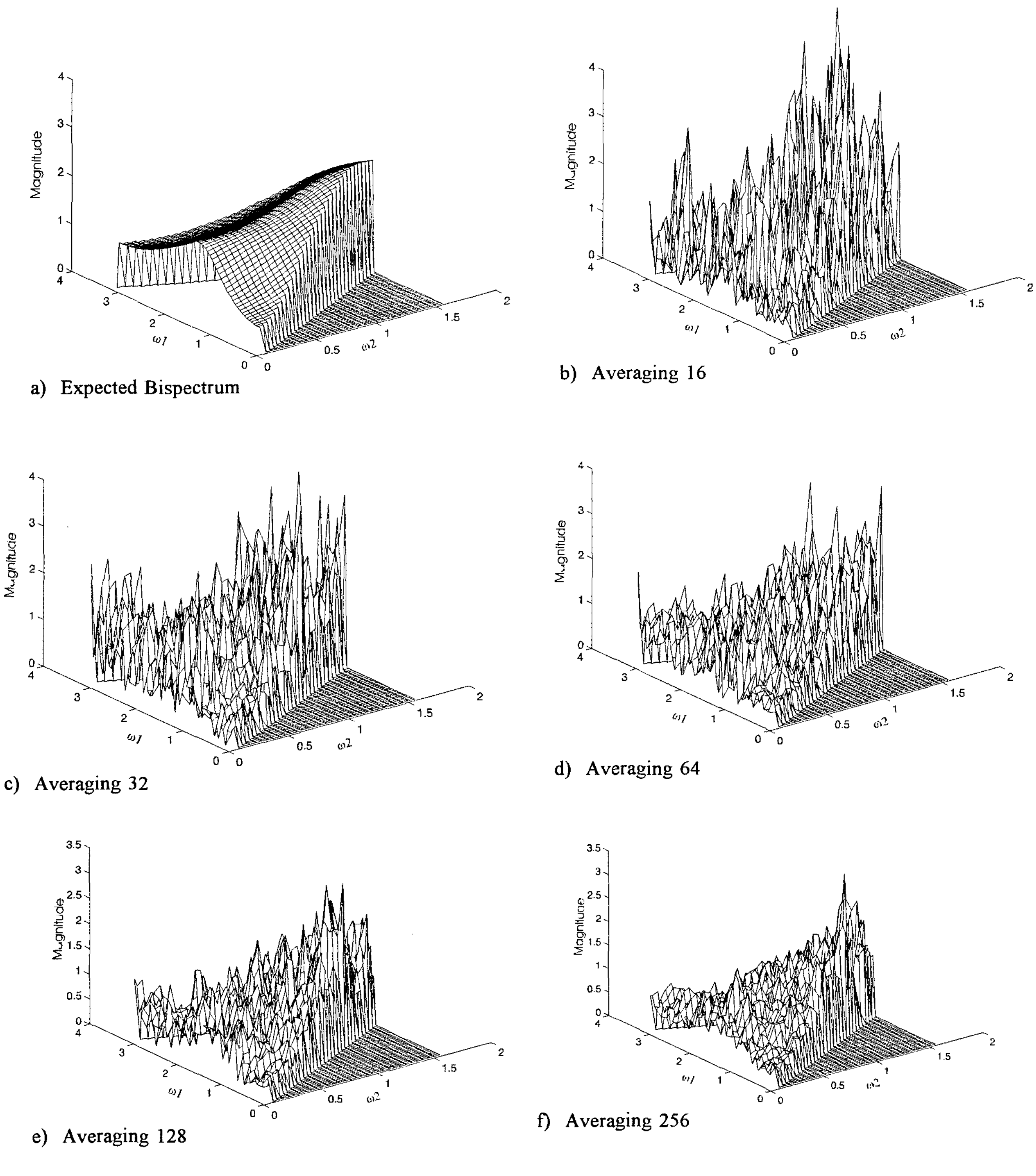


b) Non-Parametric Extimate



c) Parametric Estimate using 4 Poles

Figure 8 Bispectrum of a Fourth-Order Autoregressive Model with Non-Gaussian White Noise Input



a) Expected Bispectrum

b) Averaging 16

c) Averaging 32

d) Averaging 64

e) Averaging 128

f) Averaging 256

Figure 9 Non-Parametric Estimates of the Bispectrum of a NGWN fed AR4 process averaging over different numbers of segments more representations were averaged. The changes in the bispectrum estimate can be seen in Figure 9. The standard deviation from the theoretical bispectrum is given in Table 2. As expected, the variance decreases as the number of segments increase. The variance remains high when compared to the maximum value of the theoretical bispectrum which was 1.60.

Table 2 Variance of the Non-Parametric Estimations for Different Numbers of Segments

Number of Segments	16	32	64	128	256
Variance	0.70	0.50	0.45	0.40	0.34

3.4 Non-Gaussian Input Autoregressive Signals with Added Gaussian White Noise

Various levels of gaussian noise were added to the fourth order AR process fed by the NGWN signal. Fourth-order AR models, as well as higher order models, were used to model the resulting bispectrum. As the order increased, the noise also began to be modelled. For signal-to-noise ratios (SNR) of 8 and 16 decibels, the fourth-order model still provided a good estimate. The bispectrum began to become distorted below a SNR of 4 decibels. Each noisy signal contained the same energy, but as the gaussian noise took up more signal energy, the 3-D energy level of the signal decreased. This can be seen in the graphs (Figures 10 to 13). The measured values are shown in Table 3.

Table 3 Signal Energy under Different Noise Conditions

SNR	16 dB	8 dB	4 dB	0 dB
signal energy	0.9995	0.9995	0.9995	0.9995
3-D energy	1.4761	1.3343	0.9522	0.6150

3.5 Summary

When the bispectrum of white noise signals were estimated, the spectra were found to be relatively flat. If the white noise has a symmetric probability density function, such as the GWN,

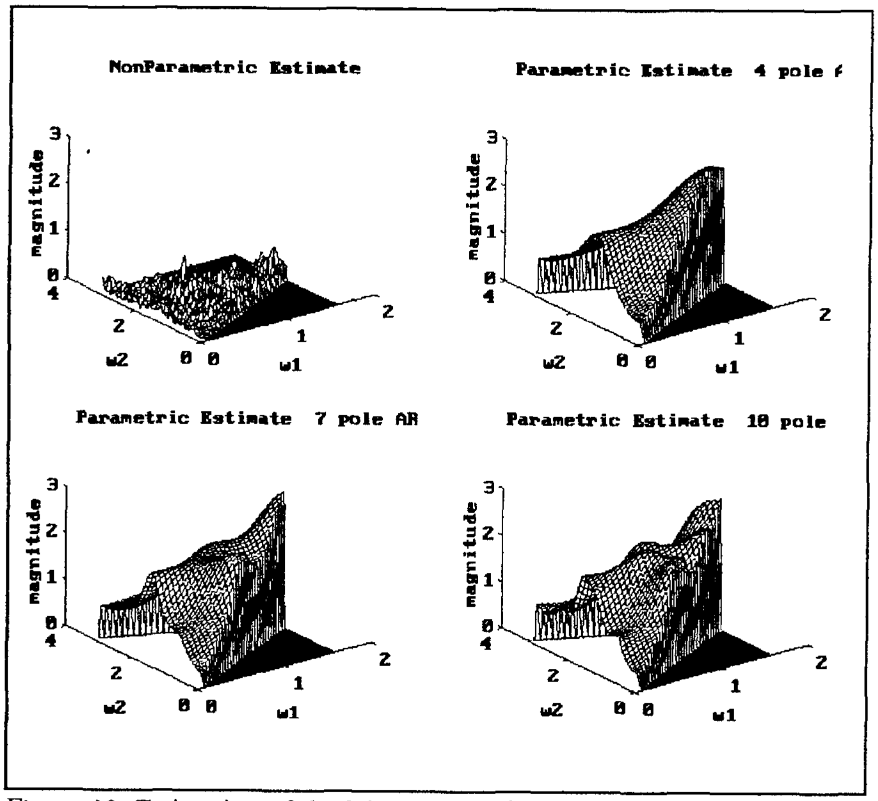


Figure 10 Estimation of the Bispectrum of an AR4 signal with a SNR of 16 dB

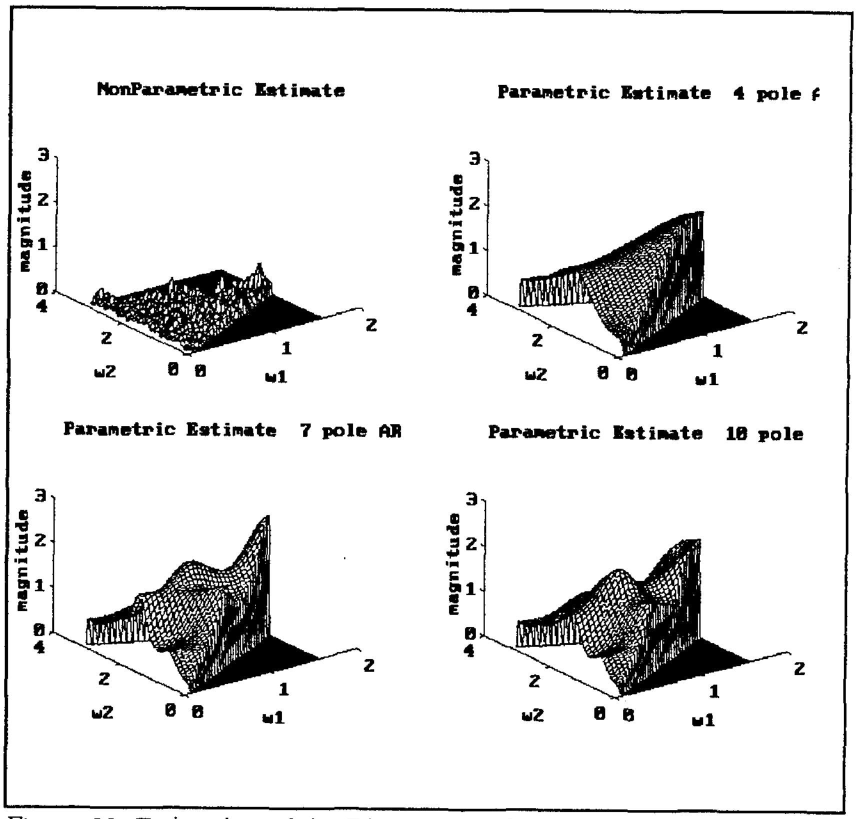


Figure 11 Estimation of the Bispectrum of an AR4 signal with a SNR of 8 dB

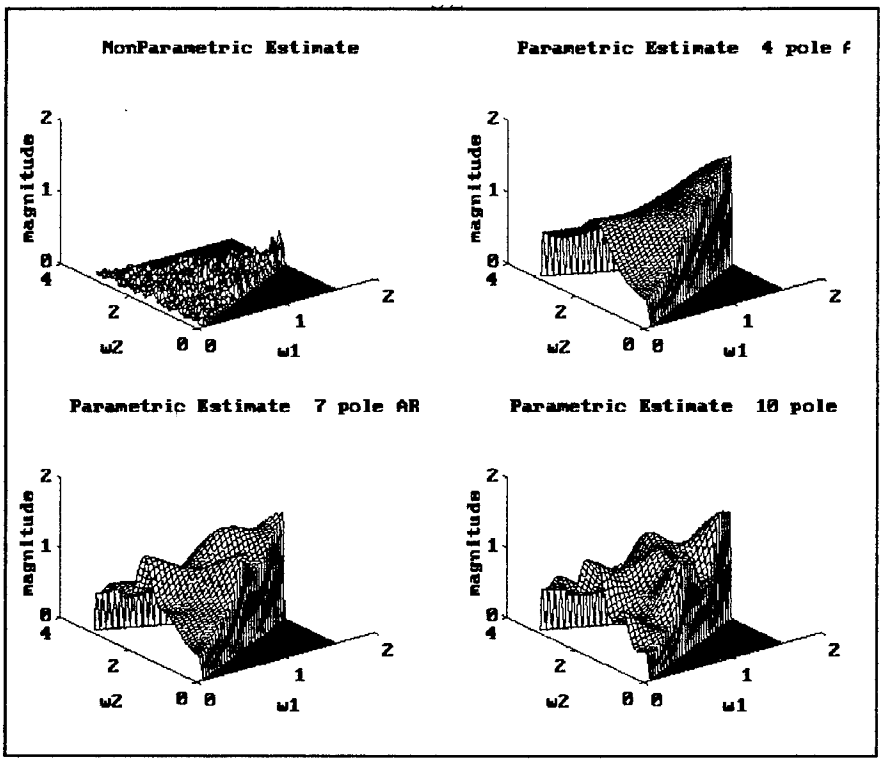


Figure 12 Estimation of the Bispectrum of an AR4 signal with a SNR of 4 dB

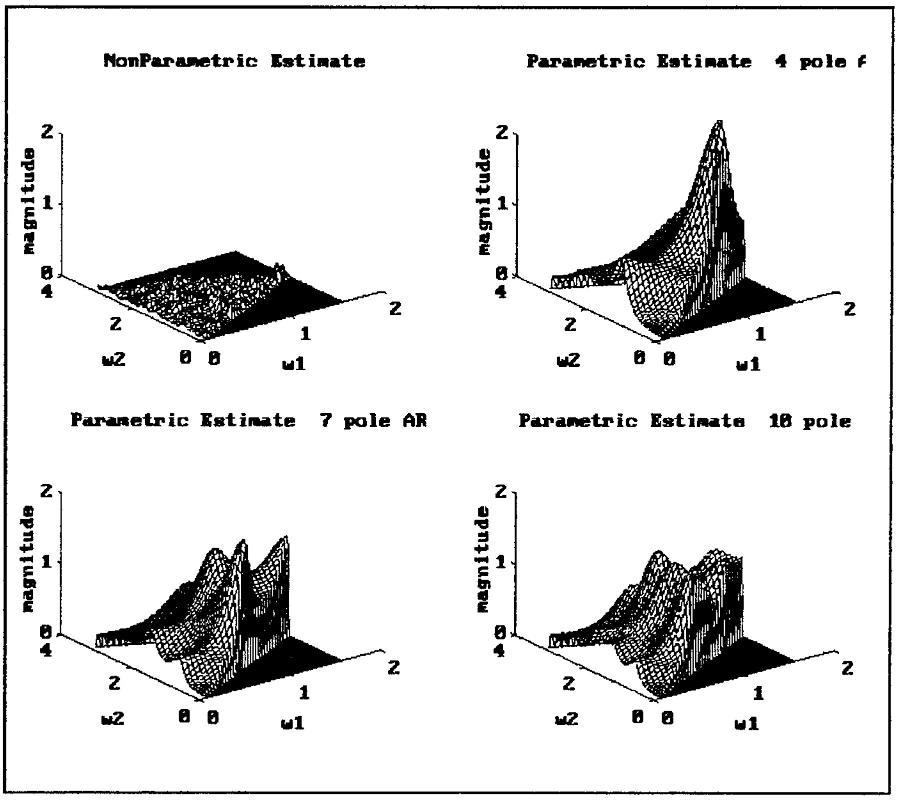


Figure 13 Estimation of the Bispectrum of an AR4 signal with a SNR of 0 dB

the value of the flat spectrum was close to zero. For a signal with a non-symmetric probability density function, such as the NGWN signal, the value was approximately constant at the 3-D energy value. These signals were fed to a linear time invariant, fourth order autoregressive process and the bispectrum of the output was estimated using both parametric and non-parametric methods. With the non-parametric method, a large number of segments needed to be averaged to reduce the variance to an acceptable level. The parametric method gave good estimates for the clean signal, but as wideband noise was added to the AR4 process, the noise began to be modelled with the signal.

IV. Application of the Bispectrum to Voice Signals

4.1 Introduction

In this chapter, the bispectrum of voice signals will be explored. The speech signals used are taken from the TIMIT data base. TIMIT is a data base of read speech that was designed to provide speech data for the acquisition of acoustic-phonetic knowledge and for the development and evaluation of automatic speech recognition systems. It resulted from efforts under sponsorship from the Defense Advanced Research Agency - Information Science and Technology Office. TIMIT contains a set of 630 speakers each speaking 10 sentences. The signals have no noise and are sampled at a rate of 16 kHz. The sentences have been converted to ascii format for use in Matlab[2]. An example of the sentence structure in the time domain is shown in Figure 14.

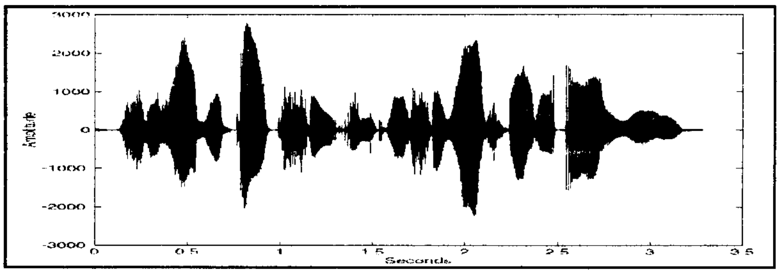


Figure 14 Time versus Amplitude for the spoken sentence "She had your dark suit in greasy wash water all year."

To study the form of the bispectrum of speech, various sounds types have been taken from the data base. Averaged bispectra of each of several sound types were estimated to determine the characteristics of speech sound. Next the effectiveness of averaging the bispectra of several segments of a signal containing additive wideband noise for noise reduction is considered. Since the signal must then be reconstructed from its bispectrum, the reconstruction process is researched. These last two areas are then combined to determine the effectiveness of using the bispectrum for speech enhancement.

4.2 Bispectrum of Different Sound Types

In this section, the bispectra of representative sounds were estimated and the results obtained by the parametric and non-parametric methods were compared. A representative result is shown in Figures 15 and 16 for the sound 'IY' as in *need*. For each sound, the period was

[2]Matlab is an interactive system and programming language for technical computation using the matrix as the basic data element.

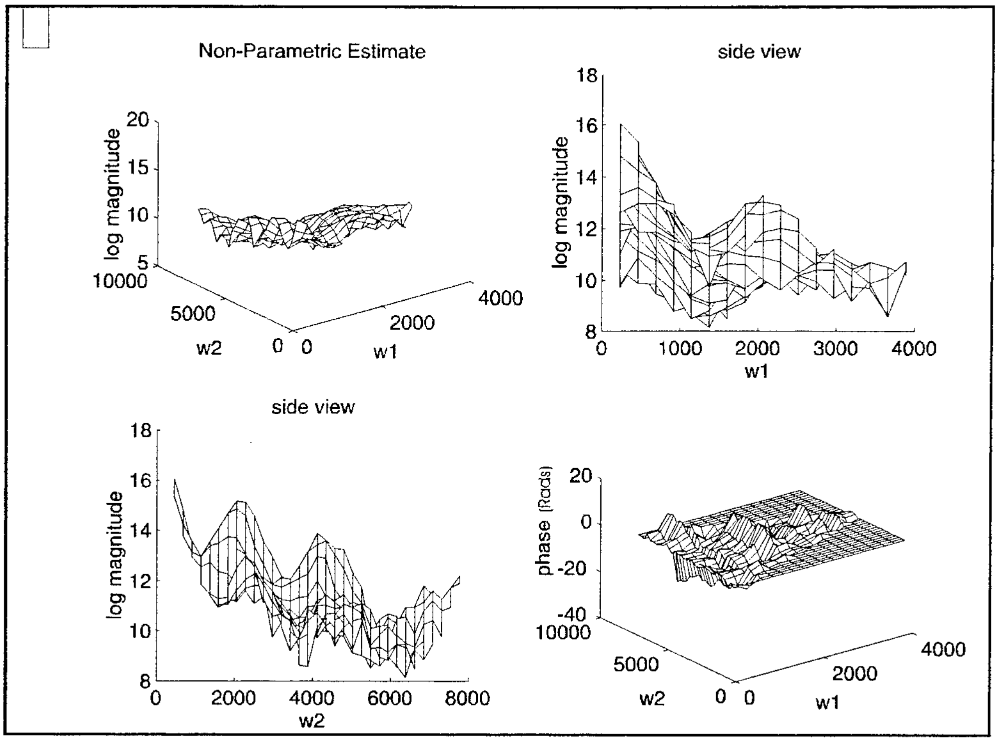


Figure 15 Non-Parametric Estimate of the Bispectrum for the Sound "IY".

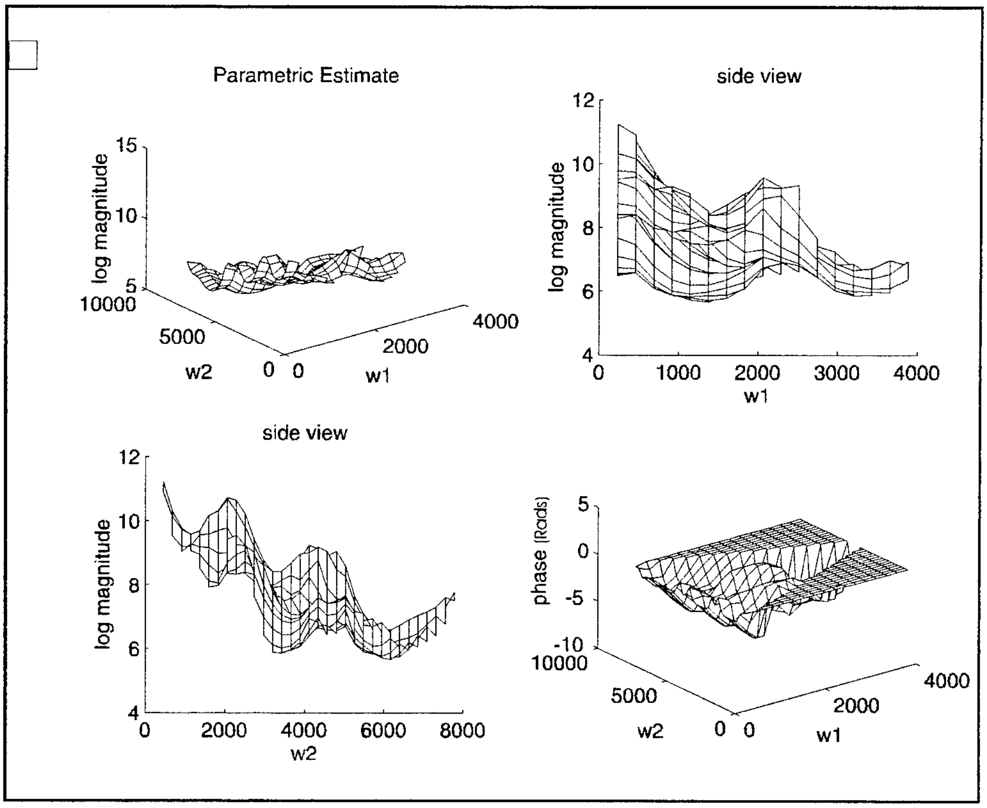


Figure 16 Parametric Estimate of the Bispectrum for the Sound "IY"

determined by comparing adjacent segments in the Mean Square Error sense. To accomplish this, the segment length was increased in increments and the difference between this segment and the ones before and after were determined at each increment. The period was taken to be the length of segment for which this difference was minimum. The bispectrum was then averaged over as many segments as were available for the duration of the sound in the sentence used. The bispectrum obtained by both the parametric and non-paramteric methods were closely matched for the voiced cases. The plots of a section of different speech sounds are shown in Figure 17 and the respective contour plots of the bispectrum are shown in Figure 18. In voiced speech, most of the energy was contained within 0 and 4 kHz, as would be expected. The fricative sound energy, such as in the sound 'SH' was spread over a wider band.

4.3 Effect of Averaging Bispectrum on Additve Wideband Noise

For this trial a fabricated stationary signal was created by repeating three periods of the sound 'IY' spoken by a female speaker. This ensured a continuous, stationary signal. The period of this fabricated signal was 74 samples. Varying levels of noise were added to this signal and the bispectrum was estimated by using the averaged bispectrum parametric and non-parametric methods. In each case, different numbers of periods were averaged to quantize the decrease in deviation due to averaging out the noise component of the bispectrum. As was discussed earlier, the expected bispectrum of gaussian white noise is zero and the bispectrum of two signals added together is the sum of the individual bispectra. Thus the expected bispectrum is that of the voice signal. Since the added gaussian white noise is ergodic, it should also have an expected time average value of zero, and the bispectrum can be averaged across time. Twenty representations of each averaged bispectrum were taken and then normalized to make the magnitudes relative, and the standard deviation throughout the bispectrum was calculated. The true bispectrum was estimated by averaging over twenty representations of the clean signal and was used as a mean value, $B_{clean}(\omega_1,\omega_2)$. Thus the variance for each frequency position was calculated by

$$var(\omega_1,\omega_2)=\sum_{i=1}^{20}\frac{(B_i(\omega_1,\omega_2)-B_{clean}(\omega_1,\omega_2))^2}{19} \qquad (52)$$

where $B_i(\omega_1,\omega_2)$ is the average estimated bispecrum of the noisy signal. The value averaged throughout the bispectrum area of interest is given in Table 4. This table does not indicate how the variance is tied to the level of the Power Spectrum, as indicated in Section 2.3.3, but it does give an indication of the affect the number of periods averaged has on the variance.

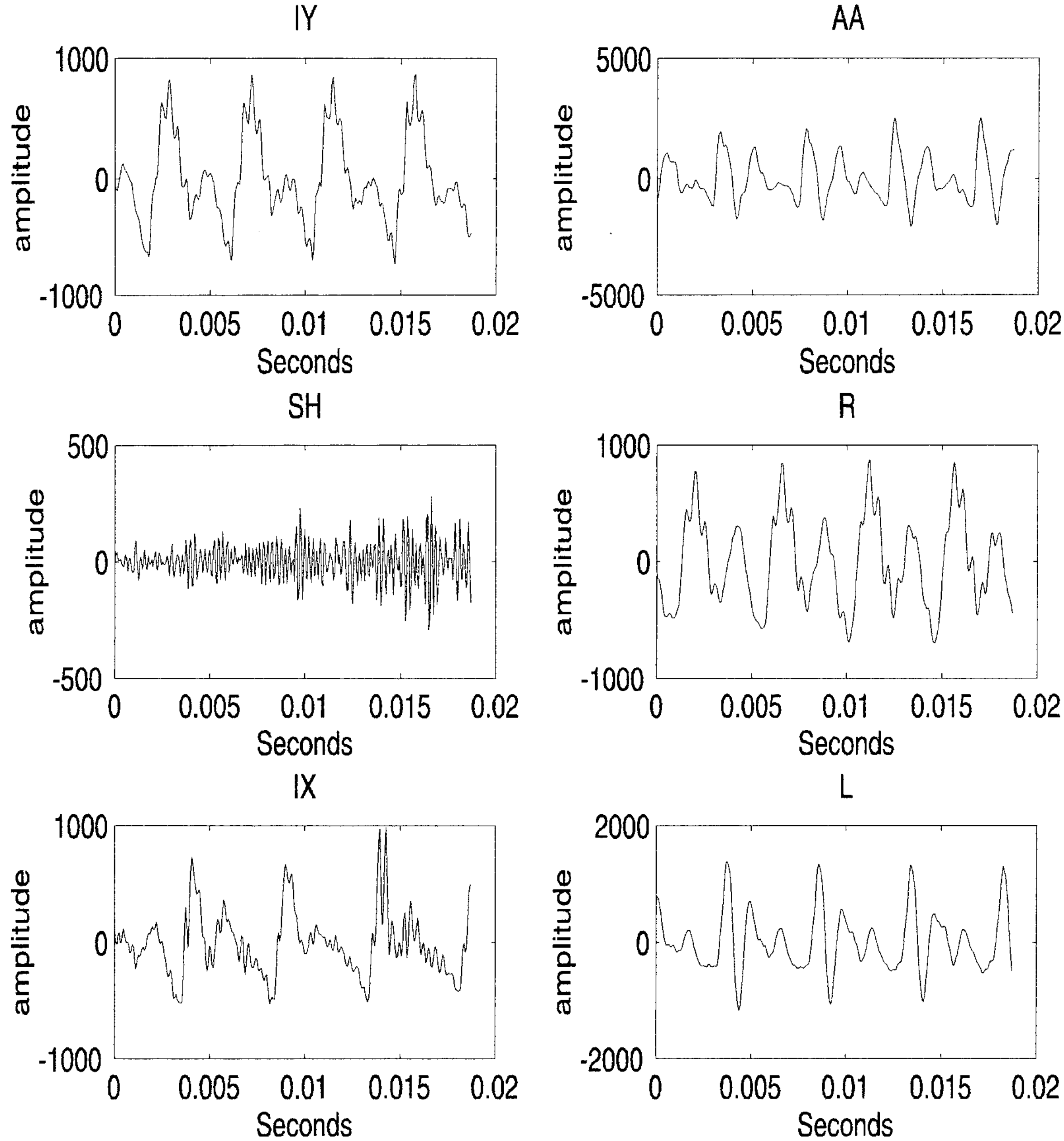


Figure 17 Sample Time Domain Plots of Various Speech Sounds

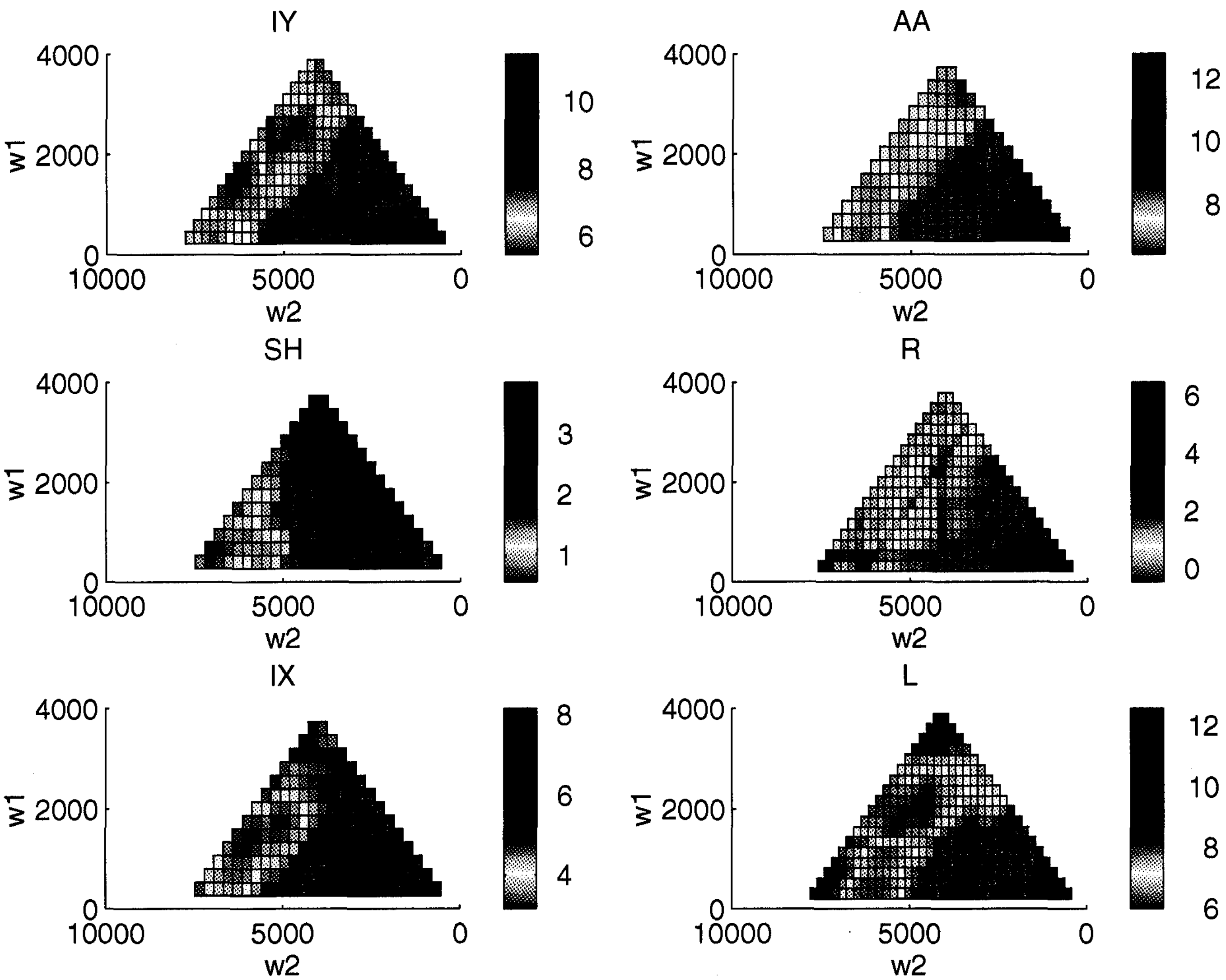


Figure 18 Contour Plots of the Bispectra corresponding to the Sounds in Figure 17

Table 4 Average Variance of Bispectrum under Noisy Conditions

	Parametric Real Part	Parametric Imaginary	Non-Para. Real Part	Non-Para. Imaginary
Signal-to-Noise Ratio = 16 dB				
1 Period	2.7	2.4	0.34	1.1
3 Periods	2.0	2.4	0.13	0.73
5 Periods	2.0	2.5	0.08	0.61
10 Periods	1.5	2.6	0.03	0.27
20 Periods	1.8	2.1	0.01	0.01
Signal-to-Noise Ratio = 8 dB				
1 Period	2.2	2.9	1.2	1.4
3 Periods	2.3	2.5	0.46	1.0
5 Periods	2.2	2.7	0.29	0.79
10 Periods	1.6	2.4	0.14	0.39
20 Periods	2.1	2.1	0.09	0.08
Signal-to-Noise Ratio = 4 dB				
1 Period	2.7	3.1	2.3	1.8
3 Periods	2.4	3.1	1.2	1.5
5 Periods	2.7	3.0	0.83	1.2
10 Periods	2.0	2.3	0.45	0.76
20 Periods	2.3	2.5	0.32	0.31
Signal-to-Noise Ratio = 0 dB				
1 Period	2.3	3.2	3.8	2.5
3 Periods	3.1	3.7	2.8	2.7
5 Periods	2.6	3.4	2.2	2.4
10 Periods	2.4	3.0	1.6	2.0
20 Periods	2.6	2.9	1.6	1.2

The variance of the Parametric estimates do not change significantly as larger number of periods are average. On the other hand, the variance does not increase significantly as higher noise levels are encountered. The statistical properties of the Parametric estimators cannot easily be predicted. The non-parametric estimate variances decrease as more periods are averaged in.

As presented in Section 2.3.3, the variances decrease as approximately a factor of the number of sections averaged. The effect this decrease in variance has on the SNR cannot be determined until the signal is reconstructed, which is covered in the next section. For the purposes of speech, a voice signal is considered stationary for a time length of 20 to 40 msec. At a sampling rate of 16 kHz, which is the rate of the signals under consideration, a single pitch period is typically 60 to 200 samples long, which is 3.8 to 12.5 msecs. Since only three periods are within the stationary time length of speech, three periods will be used in the last section to enhance speech in the presence of additive Gaussian noise.

4.4 Signal Reconstruction from the Bispectrum

In order to be able to use the bispectrum to enhance a signal, it is necessary to be able to recreate the signal from the bispectrum. The bispectrum of one period of a clean voice signal was estimated using the non-parametric method descibed in Section 2.3.1. The segment used was one period of an IY sound with a sample size of 80. The non-parametric method was used because the reconstruction method parallels the theory that the non-parametric estimation techniques are based on. The parametric estimates could give better results if similar methods as that used in Linear Predictive Coding were used. This thesis does not cover these techniques; it will concentrate on non-parametric reconstruction techniques. The magnitude reconstruction follows the algorithm laid out in Section 2.4.1. This method completely reconstructed the Fourier magnitude, as shown in Figure 19. The difficulty occurred in the reconstruction of the phase. If the Fourier phase was calculated separately from the bispectrum and left in the form given in Equation 33 without wrapping the phase to a value between +/- π, the phase can be obtained by linear methods similar to that used for magnitude recovery. This was done for the first case. The signal was fully recovered. This method is limited to cases where the phase can be stored separately. If only the bispectrum is available, the unwrapped phase values cannot be recovered, making this method unusable. In such cases, the Fourier phase can be estimated using a recursive calculation, as outlined in Section 2.4.2. The wrapped criterion is dealt with by using the exponential of the phase, which is indifferent to 2π phase shifts. The catch with this method is that the first non-DC phase cannot be determined; it is set as an arbitrary variable. The effect different values have on the signal is to shift the signal in time. In the second case considered, the value of $\psi(1)$ was set to 0. Even though the Fourier phase did not match the original in any way, as can be seen in Figure 20, the effect was to shift the original signal in time. In an attempt to reduce this shift, the value of $\psi(1)$ was set to be the same as the first non-DC Fourier phase of the original signal. The remaining phase values were fully reconstructed using the recursive method, and the orignal signal was recovered. The recovered signals for each case are shown in Figure 21.

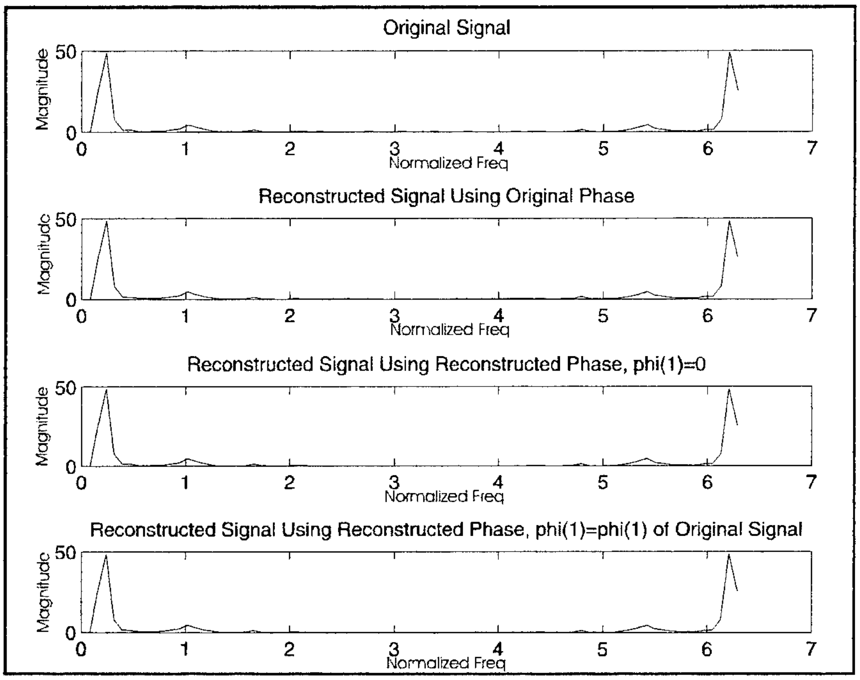


Figure 19 Fourier Magnitude of One Period of the Sound 'IY'

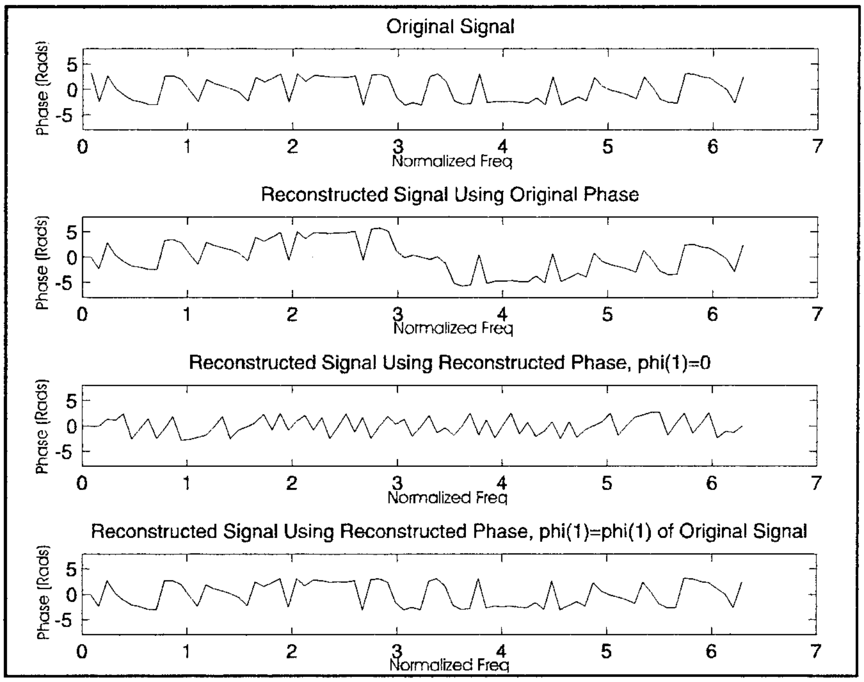


Figure 20 Fourier Phase of One Period of the Sound 'IY' .

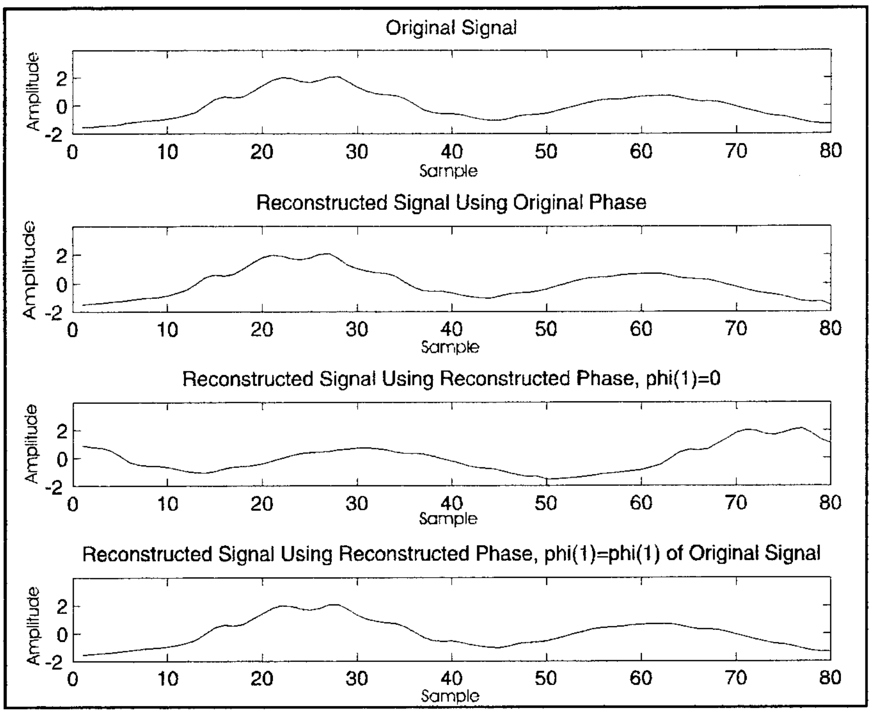


Figure 21 Original and Reconstructed Versions of One Period of the Sound "IY".

4.5 Speech Enhancement

This section presents the possible enhancement of speech that has been corrupted by additive white Gaussian noise. It compares speech that has been processed by averaging the bispectra of adjacent periods to remove the noise with averaging the time-domain speech signal on a period to period basis to remove some of the noise energy. The averaging procedure used for all cases was

$$S_j = 0.3\bar{S}_{j-1} + 0.4\bar{S}_j + 0.3\bar{S}_{j+1} \tag{53}$$

where S_j is the sequence of numbers, either in the time domain or as bispectra, corresponding to period j. This forms the weighted average of three adjacent speech periods. The emphasis is on the middle section, which will be recovered using the appropriate algorithm. The period was again determined by comparing three adjacent segments in the Mean Square Error sense. Basically, the length and position of the *j*-th frame is changed until the two adjacent segments (*j-1* and *j+1*) are the most similar to the segment in question. The three periods that were averaged are within the time requirements for stationary speech. This was only used for voiced speech segments. No processing was carried out on unvoiced, or mixed voiced/unvoiced segments.

The Fourier magnitude of the signal is recovered from the averaged bispectra using the least squares algorithm from Section 2.4.1. Since the Fourier phase cannot be recovered in a least squares method, two other methods were tried. In the first method, the Fourier phase used to reconstruct the signal is the original, noisy phase of the segment being reconstruced left in the unwrapped form. With the second method, the Fourier phase was recovered using the recursive method. The first non-DC phase was estimated by taking the first non-DC Fourier phase of the noisy signal. This works well for relatively noise-free speech. However, as the noise level increases past a SNR of 8 dB, some time shifting is evident in the recovered segments. This causes a distortion in the recovered speech signal and makes computation of the new SNR misleading. The reconstructed signals are shown in Figures 22 through 26 for various noise levels.

To determine the performance of each method, clean vowel sounds were extracted from sentences in the TIMIT data base. Different levels of zero-mean Gaussian noise were added to the clean signal to generate the corrupted signal. The corrupted signal was then processed using two of the enhancement methods; time-domain averaging and bispectrum averaging using the noisy Fourier phase. The SNR of each vowel segment was then computed. Due to the time shift in the phase reconstruction method, the SNR could not be calculated. The actual results are given in Appendix D and are summarized in Table 5.

Table 5 Signal-to-Noise Ratio After Processing (with 95% Confidence Intervals)

Initial SNR (dB)	Bispectrum Averaging with old phase	Time Domain Averaging
16	11.0643±0.4042	14.4527±0.6850
8	7.4871±0.2191	10.5997±0.3677
4	5.1237±0.0846	7.8454±0.2379
0	2.3250±0.1024	4.7482±0.1390

4.6 Summary

In this section the application of the bispectrum to speech enhancement was developed. When averaged across many representations, the parametric and non-parametric estimation techniques gave very similar results. For voiced speech most of the signal energy was in a

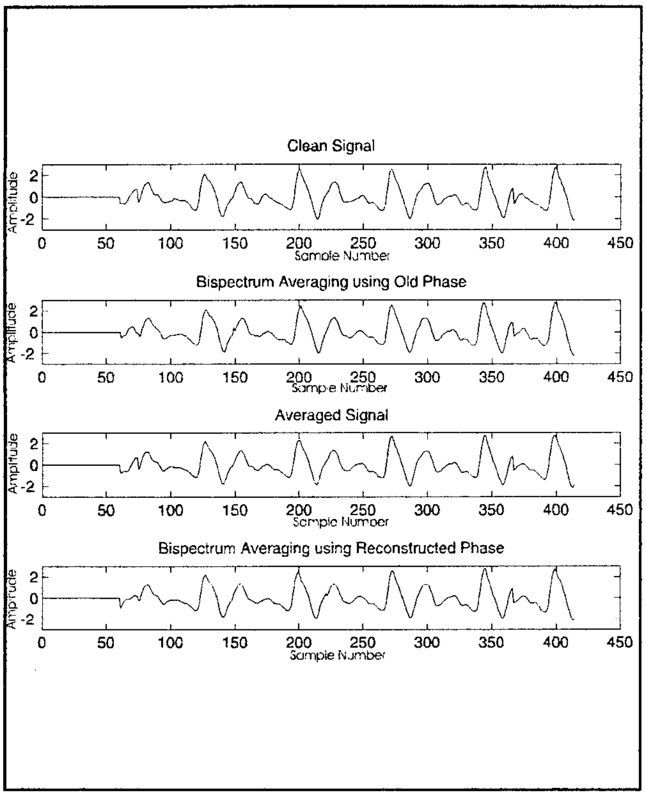


Figure 22 Signal Reconstruction - No Added Noise

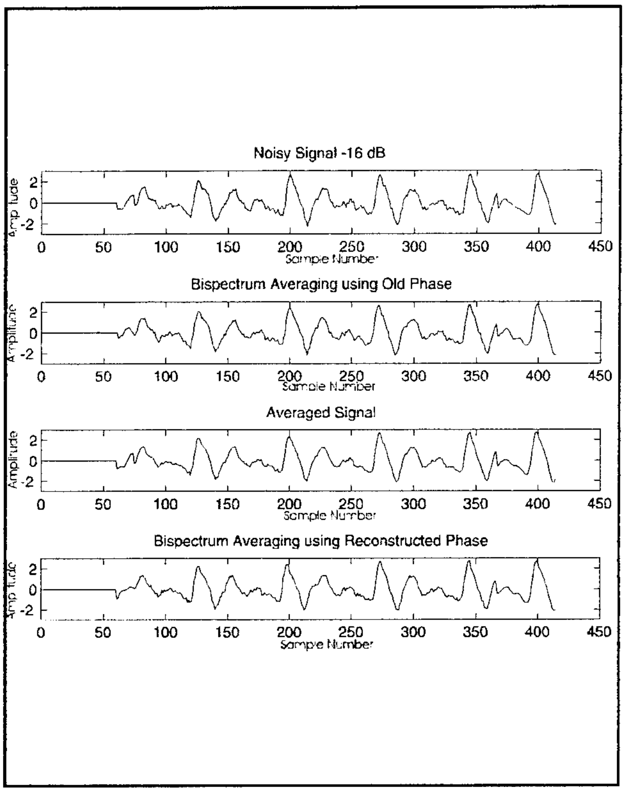


Figure 23 Signal Reconstruction - SNR of 16 dB

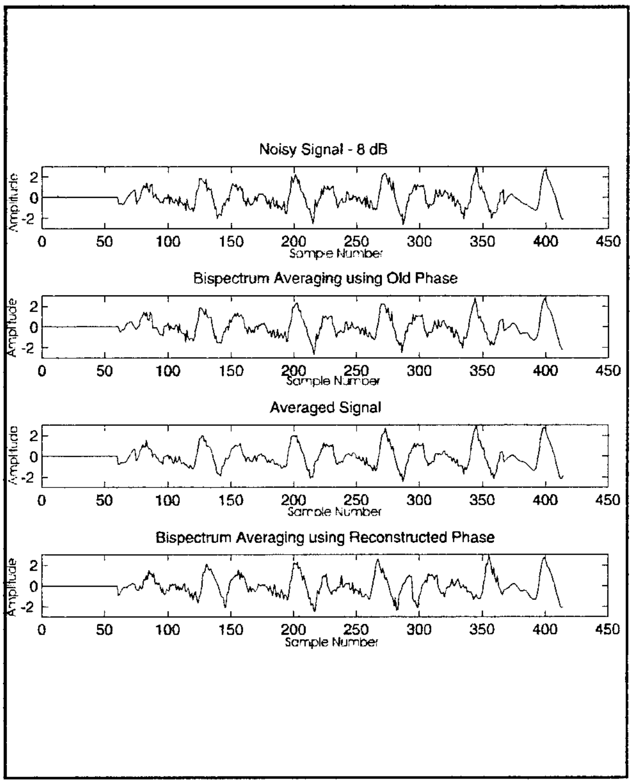


Figure 24 Signal Reconstruction - SNR of 8 dB

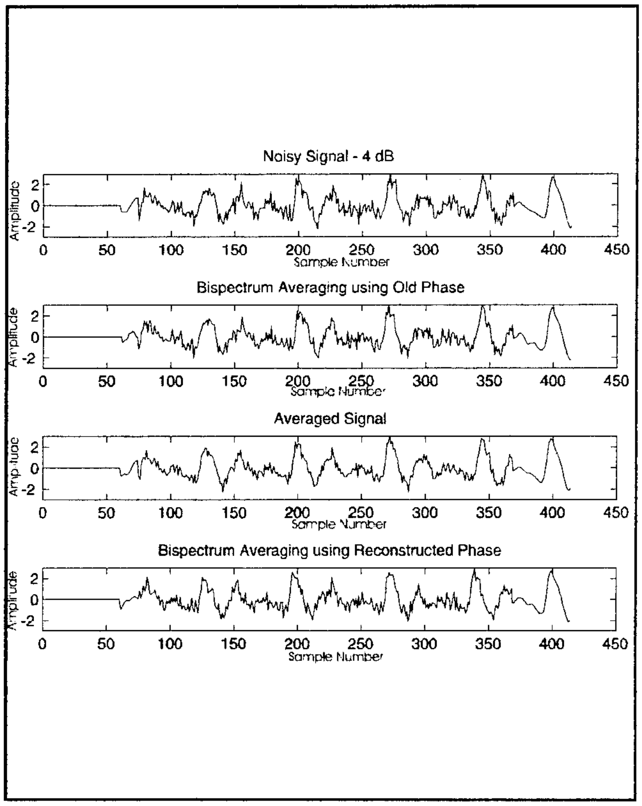


Figure 25 Signal Reconstruction - SNR of 4 dB

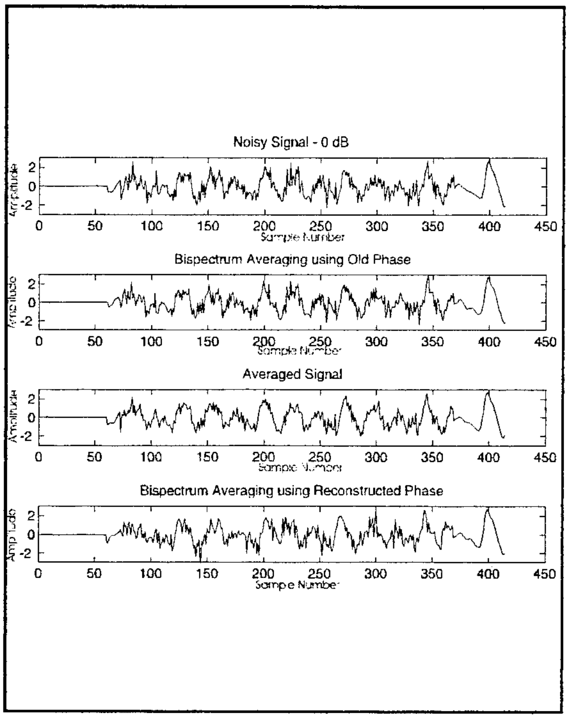


Figure 26 Signal Reconstruction - SNR of 0 dB

frequency range below 3 kHz. The signal energy was more wideband for unvoiced speech. Noise was next added to the speech and the effect of averaging the noisy speech bispectrum was researched. In parametric estimations, neither the amount of noise energy nor the number of segments averaged had much effect on the variance of the noisy bispectrum. With non-parametric estimations, the variance increased with the noise energy, but by averaging the segments, the variance could be reduced. For actual speech enhancement the non-parametric estimate was used and the bispectrum was averaged across three periods. The noisy signal was enhanced for an initial SNR of 4 dB or lower. The benefits were not significant. Better results were obtained when the signal was averaged over three periods in the time domain.

V. Conclusion

5.1 Summary

A problem of enhancing speech that was identified in the initial section of this paper is the removal of wideband random noise. The difficulty is due to the fact that broadband noise overlaps the speech signal in both time and frequency. The basis of the method of speech enhancement in this thesis is that the bispectrum has the property of being zero for a Gaussian noise signal. In addition, the bispectrum of two signals added together is the sum of the respective bispectra. Thus it was hoped that some of the noise energy could be removed from a noisy speech signal by working with the bispectra. To this end, methods of estimating the bispectrum were considered, with work carried out in both non-parametric and parametric methods. In working with the bispectrum, the signal must eventually be reconstructed. The method used in this research is to recover the Fourier magnitude and phase sequence of the signal from the bispectrum and reconstruct the signal using an inverse Discrete Fourier Transform.

Initially, the bispectrum of Gaussian and Non-Gaussian White Noise signals were estimated and their characteristics studied. The expected value of the bispectrum of a Gaussian White noise signal is zero. Results of the estimated bispectrum were given to be close to zero (in the order of 10^{-6}). The expected bispectrum of the Non-Gaussian White noise was to be flat at the value of what this thesis termed as the 3-D energy estimated by the averaged sum of the cube of each value in a sequence. As more segments were averaged, the variance of the result decreased. This is not surprising, since the sequence is a random process, and its only the expected bispectrum that is flat. In a second part, the white noise signals were used as inputs to a fourth -order Auto Regressive process. Since the bispectrum of the output process can be calculated as a product of the input bispectrum, the expected output of the process fed by Gaussian noise was zero. The value was actually larger than zero, but much smaller than that provided by the non-Gaussian input.

In the final section of this thesis, the bispectrum was applied to noice signals. First the bispectra of different sound types were estimated. For voiced speech, the energy was concentrated in a the frequency band less than 4 kHz. In unvoiced sounds, such as the sound "SH", the energy was more wideband. Each sound did appear to have a characteristic bispectrum, though similarities would make it hard to identify individual sounds. Next the effect of averaging various segments of an estimated bispectrum has on additive wideband noise was investigated. Since the expected value of the bispectrum of wideband noise is zero, averaging the bispectra of a signal to which noise has been added should reduce the effect of the noise. This was found for the case of the bispectra estimated using non-parametric methods. The variance of parametric methods is not fully understood. In this research, it was found that neither the noise level, nor the number of segments averaged had any significant effects on the variance of the bispectra.

Finally, the use of averaged bispecta for enhancing noisy speech signals was considered. It was found that for initial signal-to-noise ratios of less than 6 dB, the overall output signal-to-noise ratio was enhanced by about 1 to 2 dB. Better results, however, were obtained by strictly averaging the time domain signal with much less computational effort.

5.2 Conclusion

The theory of the bispectrum was borne out through the experiments and tests carried out in this research. It is in the applicability of this theory to enhance voice signals that is in doubt. Since any additive noise will be a random process, it is only the expected value of the bispectrum that is zero. With speech signals, the signal is only considered stationary for a period of 20 to 40 milliseconds. This only allows the bispectra of three to four periods to be used for averaging within that stationary time. This does not allow a significant amount of the noise energy to be removed through the averaging process. Some energy was removed, but classical methods are just as effective.

5.3 Recommendations for Future Consideration

Due to the nature of the reconstruction techniques used, this thesis was restricted to the use of the non-parametric estimates. The parametric estimates, though higher in variance, did not seem to be as affected by the noise. Since the coefficients obtained by the parametric estimation of the bispectrum are theoretically the same as those obtained with Linear Predictive Coding methods, which use second-order methods, they should be effective with Linear Predictive Coding algorithms. The classical problem of second-order Linear Predictive techniques is that they are highly inaccurate in noisy conditions. Parametric bispectrum estimates should give more accurate coefficient estimates.

Appendix A: General Theory of the Third-Order Spectra

A.1 General Moment and Cumulant Theory[3]

Given a set of n real random variables $\{x_1, x_2, \ldots, x_n\}$, their joint cumulants of order $r = k_1 + k_2 + \ldots + k_n$ are defined as

$$c_{k_1 \cdots k_n} \triangleq (-j)^r \frac{\delta^r \ln \phi(\omega_1, \omega_2, \ldots, \omega_n)}{\delta \omega_1^{k_1} \delta \omega_2^{k_2} \ldots \delta \omega_n^{k_n}} \Big|_{\omega_1 = \omega_2 = \ldots = \omega_n = 0}, \tag{54}$$

where

$$\phi(\omega_1, \omega_2, \ldots \omega_n) = E[\exp j(\omega_1 + \ldots + \omega_n)]$$

is their *joint characteristic function*. Note that the joint moments of order r of the same set of random variables are given by

$$\begin{aligned} m_{k_1 \ldots k_n} &\triangleq E[x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}] \\ &= (-j)^r \frac{\delta^r \phi(\omega_1, \omega_2, \ldots, \omega_n)}{\delta \omega_1^{k_1} \ldots \delta \omega_n^{k_n}} \Big|_{\omega_1 = \ldots = \omega_n = 0.} \end{aligned} \tag{56}$$

Hence, the *joint cumulants* can be expressed in terms of the joint moments of the set of random variables. For example, the moments

$$m_1 = E[x_1] \qquad m_2 = E[x_1^2]$$
$$m_3 = E[x_1^3]$$

of the random variable $\{x_1\}$ are related to its cumulants by

$c_1 = m_1$

$c_2 = m_2 - m_1^2$

$c_3 = m_3 - 3m_2 m_1 + 2m_1^3$

$c_4 = m_4 - 4m_3 m_1 - 3m_2^2 + 12m_2 m_1^2 - 6m_1^4.$

---

[3] The background for this theory is taken from Nikias [12]

Hence the computation of the nth order cumulant requires knowledge of all its moments from the first to the nth order.

If $\{x(k)\}$, $k=0,\pm 1,\pm 2,\ldots$ is a real, strictly stationary random process and its moments up to nth order exist, then the moment will depend only on the time differences τ_1, τ_2, ..., τ_{k-1} and therefore we can write

$$m_n(\tau_1,\tau_2,\ldots,\tau_{n-1})=E\{x(k)x(k+\tau_1)\ldots x(k+\tau_{n-1})\}.$$

For orders $n=1,2,3$ the cumulants $c_n(\tau_1,\ldots,\tau_{n-1})$ are related to the moments of $\{x(k)\}$ as follows:

$$c_1=m_1=E\{x(k)\}$$
$$c_2(\tau_1)=m_2(\tau_1)-m_1^2$$
$$c_3(\tau_1,\tau_2)=m_3(\tau_1,\tau_2)-m_1[m_2(\tau_1)+m_2(\tau_2)+m_2(\tau_2-\tau_1)]+2m_1^3.$$

If the process is zero mean ($m_1=0$), it follows that the second and third order cumulants are identical to the second- and third-order moments, respectively. Thus

$$c_3(\tau_1,\tau_2)=m_3(\tau_1,\tau_2)=E[x(k)x(k+\tau_1)x(k=\tau_2)] \tag{58}$$

which is given as Equation 1.

A.2 Higher Order Spectra

Higher order spectra of signals are usually defined in terms of cumulants and not moments. The nth-order spectrum $B_n(\omega_1,\ldots,\omega_{n-1})$ of $\{x(l)\}$ is defined as the (n-1) dimensional Fourier transform of the nth-order cumulant sequence, that is,

$$\begin{aligned}B_n(\omega_1,\omega_2,\ldots,\omega_{n-1})=&\frac{1}{(2\pi)^{n-1}}\sum_{\tau_1=-\infty}\cdots\sum_{\tau_{n-1}=-\infty}\\&\cdot c_n(\tau_1,\tau_2,\ldots,\tau_{n-1})\\&\cdot\exp\left[-j(\omega_1\tau_1+\omega_2\tau_2+\ldots+\omega_{n-1}\tau_{n-1})\right]\\&|\omega_i|\le\pi \ \textit{for}\ i=1,2,\ldots,n-1\\&|\omega_1+\omega_2+\ldots+\omega_{n-1}|\le\pi\end{aligned} \tag{59}$$

In general, $B_n(\omega_1,\ldots,\omega_{n-1})$ is complex for $n>2$, so it has magnitude and phase. The cumulant spectrum is also a periodic function with period 2π.

The power spectrum and bispectrum are special cases of the nth-order cumulant spectrum defined by the equation above.

Power spectrum: n=2

$$P_n=\frac{1}{2\pi}\sum_{\tau=-\infty}^{+\infty} c_2(\tau)\exp[-j(\omega\tau)] \qquad |\omega|\le\pi, \tag{60}$$

where $c_2(\tau)$ is the covariance sequence of $\{x(k)\}$.

Bispectrum: n=3

$$C_3(\omega_1,\omega_2)=\frac{1}{(2\pi)^2}\sum_{\tau_1=-\infty}^{+\infty}\sum_{\tau_2=-\infty}^{+\infty} c_3(\tau_1,\tau_2)\exp[-j(\omega_1\tau_1+\omega_2\tau_2)] \qquad |\omega_1|\le\pi,\ |\omega_2|\le\pi,\ |\omega_1+\omega_2|\le\pi, \tag{61}$$

where $c_3(\tau_1,\tau_2)$ is the third-order cumulant sequence of $\{x(k)]$. In the paper $B(\omega_1,\omega_2)$ will be used for $C_3(\omega_1,\omega_2)$. The above equation is given as Equation 2.

A.3 Symmetries

The third order cumulant

$$c_3(t_1,t_2,t_3)=E[x(t_1)x(t_2)x(t_3)]$$

is invariant to the six permutations of the numbers t_1, t_2, and t_3. For stationary processes $t_2-t_1=\tau_1$, $t_3-t_1=\tau_2$, and $t_3-t_2=\tau_2-\tau_1$. This yields the identities

	Non-Stationary	Stationary	Spectrum
1	t_1,t_2,t_3	τ_1,τ_2	ω_1,ω_2
2	t_2,t_1,t_3	$-\tau_1,\tau_2-\tau_1$	$-\omega_1-\omega_2,\omega_2$
3	t_3,t_1,t_2	$-\tau_2,\tau_1-\tau_2$	$\omega_2,-\omega_1-\omega_2$
4	t_3,t_2,t_1	$\tau_1-\tau_2,-\tau_2$	$\omega_1,-\omega_1-\omega_2$
5	t_2,t_3,t_1	$\tau_2-\tau_1,-\tau_1$	$-\omega_1-\omega_2,\omega_1$
6	t_1,t_3,t_2	τ_2,τ_1	ω_2,ω_1

Since the cumulant function is real, the symmetry $B(-\omega_1,-\omega_2)=B^*(\omega_1,\omega_2)$ also exists. Thus, if we know $B(\omega_1,\omega_2)$ in any one of the twelve regions of Figure 19 we can determine it everywhere.

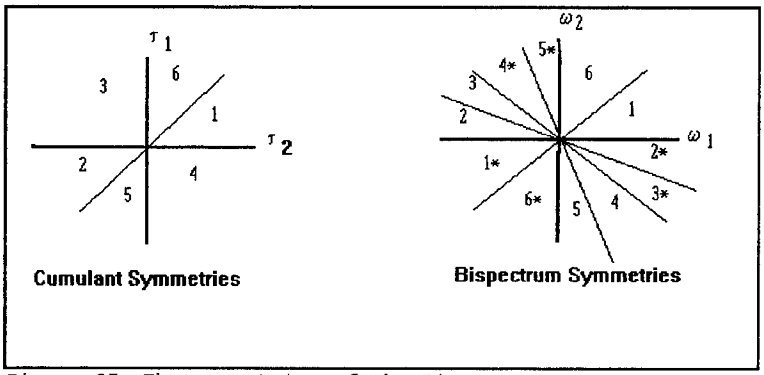


Figure 27 The symmetries of the Bispectrum

Appendix B: *Properties of the Bispectrum*

B.1 Additive Property of Independant Cummulants

Let $z=x+y=(x_1+y_1,...,x_n+y_n)$. For a zero mean process

$$\begin{aligned} c_3^z(\tau_1,\tau_2) &= E[z(k)z(k+\tau_1)z(k+\tau_2)] \\ &= E[(x(k)+y(k))(x(k+\tau_1)+y(k+\tau_1))(x(k+\tau_2)+y(k+\tau_2))] \\ &= E[x(k)x(k+\tau_1)x(k+\tau_2)+E[y(k)y(k+\tau_1)y(k+\tau_2)] \\ &= c_3^x(\tau_1,\tau_2)+c_3^y(\tau_1,\tau_2) \end{aligned} \tag{63}$$

since all cross terms when multiplying the expected values of x and y go to zero because x and y are independantly distributed.

B.2 Symmetric Probability Density Functions

This property will be demonstrated for a random gaussian process, since this is the signal type this thesis utilizes. By definition:

$$E[x_1x_2x_3]=\iiint x_1x_2x_3f(x_1,x_2,x_3;t_1,t_2,t_3)dx_1dx_2dx_3 \tag{64}$$

Since, x is from a random gaussian distribution, x_1,x_2,x_3 are iid and

$$\begin{aligned} E[x_1,x_2,x_3] &= \iiint x_1x_2x_3\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x_1-m)^2}{2\theta^2}}\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x_2-m)^2}{2\theta^2}}\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x_3-m)^2}{2\theta^2}}dx_1dx_2dx_3 \\ &= \int x_1\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x_1-m)^2}{2\theta^2}}dx_1\int x_2\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x_2-m)^2}{2\theta^2}}dx_2\int x_3\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x_3-m)^2}{2\theta^2}}dx_3 \end{aligned} \tag{65}$$

Taking each integral separately, ie

$$\begin{gathered} \int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-m)^2}{2\theta^2}}dx \quad \textit{but the expression} \\ x\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-m)^2}{2\theta^2}} \end{gathered} \tag{66}$$

is odd since the probability density function is even, therefore the integral, and hence the expectation becomes zero.

B.3 Linear Systems

Assume $x\{k\}$ is the input to a linear-time-tinvariant (LTI) system described by

$$y(k)=\sum_{i=-\infty}^{+\infty} h(k-i)x(i)$$

where $h(k)$ is the impulse response of the system. The autocorrelation of $y(k)$ can be obtained by [aa12, 311]

$$R_{yy}(t_1,t_2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} R_{xx}(t_1-\alpha,t_2-\beta)h(\alpha)h(\beta)d\alpha d\beta. \tag{68}$$

Extending this to the third order moments

$$R_{yyy}(t,t-\tau_1,t-\tau_2)=E[y(t)y(t-\tau_1)y(t-\tau_2)] \tag{69}$$

which, in terms of the third order moment R_{xxx} of x(t), becomes

$$R_{yyy}(t,t-\tau_1,t-\tau_2)=\iiint_{-\infty}^{\infty} R_{xxx}(t-\alpha,t+\tau_2-\beta,t+\tau_3-\gamma)h(\alpha)h(\beta)h(\gamma)d\alpha d\beta d\gamma. \tag{70}$$

For stationary processes,

$$R_{xxx}(t,t-\tau_1,t-\tau_2)=R_{xxx}(\tau_1,\tau_2);$$

hence

$$R_{yyy}(\tau_1,\tau_2)=\iiint_{-\infty}^{\infty} R_{xxx}(\tau_1+\alpha-\beta,\tau_2+\alpha-\gamma)h(\alpha)h(\beta)h(\gamma)d\alpha d\beta d\gamma. \tag{72}$$

Taking transformations of both sides of this equation

$$\begin{aligned} S_{yyy}(\omega_1,\omega_2) &= \iiint_{-\infty}^{\infty}\iint_{-\infty}^{\infty} R_{xxx}(\tau_1+\alpha-\beta,\tau_2+\alpha-\gamma)e^{-j(\omega_1\tau_1+\omega_2\tau_2)}d\tau_1 d\tau_2 h(\alpha)h(\beta)h(\gamma)d\alpha d\beta d\gamma \\ &= \iiint_{-\infty}^{\infty} e^{-j(\omega_1(\alpha-\beta)+\omega_2(\alpha-\gamma))}\iint_{-\infty}^{\infty} R_{xxx}(\tau_1+\alpha-\beta,\tau_2+\alpha-\gamma) \\ &\quad * e^{-j(\omega_1(\tau_1+\alpha-\beta)+\omega_2(\tau_2+\alpha-\gamma))}d\tau_1 d\tau_2 h(\alpha)h(\beta)h(\gamma)d\alpha d\beta d\gamma \\ &= S_{xxx}(\omega_1,\omega_2)\iiint_{-\infty}^{\infty} e^{-j(\omega_1(\alpha-\beta)+\omega_2(\alpha-\gamma))}h(\alpha)h(\beta)h(\gamma)d\alpha d\beta d\gamma \\ &= S_{xxx}(\omega_1,\omega_2)\iiint_{-\infty}^{\infty} h(\beta)e^{-j\omega_1\beta}h(\gamma)e^{-j\omega_2\gamma}h(\alpha)e^{j(\omega_1+\omega_2)\alpha}d\alpha d\beta d\gamma. \end{aligned} \tag{73}$$

And thus

$$S_{yyy}(\omega_1, \omega_2) = S_{xxx}(\omega_1, \omega_2) H(\omega_1) H(\omega_2) H^*(\omega_1 + \omega_2). \tag{74}$$

B.4 Relationship between the Bispectrum and the Fourier Transform

By definition, the third-order cumulant of discrete, zero-mean, ergotic systems can be given by:

$$c_3(m,n) = E[x(k)x(k+m)x(k+n)] = \frac{1}{K}\sum_{k=-\infty}^{\infty} x(k)x(k+m)x(k+n) \tag{75}$$

The bispectrum is the two dimensional Fourier Transform of the above expression:

$$\begin{aligned} B(\omega_1, \omega_2) &= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{1}{K}\sum_{k=-\infty}^{\infty} x(k)x(k+m)x(k+n)e^{-j(\omega_1 m+\omega_2 n)} \\ &= \frac{1}{K}\sum_{k}\sum_{m}\sum_{n} x(k)e^{j(\omega_1+\omega_2)k}x(k+m)e^{-j\omega_1(k+m)}x(k+n)e^{-j\omega_2(k+n)} \\ \text{since } & e^{-j(\omega_1 m+\omega_2 n)} = e^{j(\omega_1+\omega_2)k - j\omega_1(k+m) - j\omega_2(k+n)}. \end{aligned} \tag{76}$$

Comparing this with the definition of the Fourier Transform,

$$X(\omega) = \sum_{\tau} x(\tau)e^{-j\omega\tau} \tag{77}$$

holding k constant and transforming variables, the expression for $B(\omega_1, \omega_2)$ can be shown to be

$$B(\omega_1, \omega_2) = \frac{1}{K}X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2). \tag{78}$$

A similar case can be made for periodic signals with the result

$$B(\omega_1, \omega_2) = \frac{1}{N}X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2) \tag{79}$$

where N is the period length.

Appendix C: ***Derivation of the AR Triple Correlation***[4]

The AR process can be described by

$$X(k)+\sum_{i=1}^{p}a_iX(k-i)=W(k) \tag{80}$$

where W(k) is the input sequence. To calculate the triple correlations, the expected value of the multiplication of the sequence taken at three different points of time is determined.

$$c_3(\tau_1,\tau_2)=E[X(k)X(k+\tau_1)X(k+\tau_2) \tag{81}$$

To do this, three cases are considered.

C.1 Case 1 τ_1 *and* τ_2 *are both positive.*

Multiplying both sides of (C-1) by $X(k+\tau_1)X(k+\tau_2)$ and taking expectations we get

$$\begin{gathered}E[X(k)X(k+\tau_1)X(k+\tau_2)]+\sum_{i=1}^{p}a_iE[X(k-i)X(k+\tau_1)X(k+\tau_2)]\\=E[X(k+\tau_1)X(k+\tau_2)W(k)]=0,\end{gathered} \tag{82}$$

i.e.,

$$\begin{gathered}c(\tau_1,\tau_2)+\sum_{i=1}^{p}a_ic(i+\tau_1,i+\tau_2)=0\\ \text{for}\ \ \tau_1,\tau_2>0.\end{gathered} \tag{83}$$

C.2 Case 2 $\tau_1=0$ *and* $\tau_2>0$.

Setting τ_1=0 in (C-3) we get

$$\begin{gathered}E[X^2(k)X(k+\tau_2)]+\sum_{i=1}^{p}a_iE[X(k-i)X(k)X(k+\tau_2)]\\=E[X(k)X(k+\tau_2)W(k)].\end{gathered} \tag{84}$$

Now multiplying both sides of (C-1) by $X(k+\tau_2)W(k)$ and taking expectations we get

[4] Taken from Rughuveer and Nikias [17:47].

$$E[X(k)X(k+\tau_2)W(k)]+\sum_{i=1}^{p}a_iE[X(k-i)X(k+\tau_2)W(k)] \\ =E[X(k+\tau_2)W^2(k)] \tag{85}$$

Substituting the result of (C-6) into (C-5) we get

$$c(\tau_1,\tau_2)+\sum_{i=1}^{p}a_ic(i+\tau_1,i+\tau_2)=0 \\ for\ \ \tau_1=0,\tau_2>0. \tag{86}$$

From the symmetry of third moment sequences it follows that the result holds when $\tau_1>0$ and $\tau_2=0$.

B.3 Case 3 $\tau_1=\tau_2=0$.

Setting $\tau_2=0$ in (C-5) we get

$$E[X^3(k)]+\sum_{i=1}^{p}a_iE[X(k-i)X^2(k)] \\ =E[X^2(k)W(k)] \tag{87}$$

Multiplying both sides of (C-1) by X(n)W(n), taking expectations and then using (B-6) we get

$$E[X^2(k)W(k)]+\sum_{i=1}^{p}a_iE[X(k-i)X(k)W(k)] \\ =E[X(k+\tau_2)W^2(k)] \\ i.e. \\ E[X^2(k)W(k)]=E[X(k)W^2(k)]. \tag{88}$$

However, multiplying (C-1) by $W^2(k)$ and taking expectations we get

$$E[X(k)W^2(k)]+\sum_{i=1}^{p}a_iE[X(k-i)W^2(k)] \\ =E[W^3(k)]. \tag{89}$$

From (C-8), (C-9) and (C-10) we have

$$c(\tau_1,\tau_2)+\sum_{i=1}^{p}a_ic(i+\tau_1,i+\tau_2)=\beta \\ for\ \ \tau_1=0,\tau_2=0. \tag{90}$$

Appendix D: ***Speech Signal Enhancement Results***

Initial SNR = 0 dB

Bispectrum Averaging using Old Phase					Time Domain Averaging				
Sound	Resulting SNR				Sound	Resulting SNR			
AA	2.2692	2.5870			AA	5.3538	4.9436		
AO	2.6139	2.3147			AO	5.0266	3.6643		
EH	2.6112				EH	5.4623			
IH	2.4313				IH	3.3667			
IY	2.2930	2.5871	2.4809	2.4360	IY	4.8638	5.1396	5.0065	5.0567
IX	0.7486				IX	3.9066			
UX	2.5276				UX	5.1872			

Initial SNR = 4 dB

Bispectrum Averaging using Old Phase					Time Domain Averaging				
Sound	Resulting SNR				Sound	Resulting SNR			
AA	5.4336	4.9436			AA	7.9237	7.7496		
AO	5.0623	4.4538			AO	8.9828	6.4338		
EH	5.3380				EH	8.5395			
IH	4.5643				IH	5.5177			
IY	5.3525	5.5088	5.4193	5.4384	IY	8.6142	8.6945	7.8770	9.0307
IX	4.4976				IX	6.1234			
UX	5.5720				UX	8.6582			

Initial SNR = 8 dB

Bispectrum Averaging using Old Phase					Time Domain Averaging				
Sound	Resulting SNR				Sound	Resulting SNR			
AA	8.0782	7.7856			AA	11.022	10.391		
AO	8.0875	6.2840			AO	12.776	8.3327		
EH	7.7923				EH	11.607			
IH	6.3523				IH	6.7365			
IY	7.6735	8.6755	7.2445	8.0978	IY	10.772	12.127	10.692	11.917
IX	5.0212				IX	8.5492			
UX	8.7526				UX	12.273			

Initial SNR = 16 dB

Bispectrum Averaging using Old Phase					Time Domain Averaging				
Sound	Resulting SNR				Sound	Resulting SNR			
AA	12.047	12.746			AA	15.064	15.594		
AO	12.050	9.2860			AO	18.914	10.451		
EH	11.027				EH	15.632			
IH	8.0869				IH	7.9971			
IY	12.180	12.400	9.2565	13.207	IY	14.484	16.813	13.477	17.346
IX	7.4266				IX	9.7942			
UX	13.058				UX	17.865			

Bibliography

[1] Bartelt, H., A.W. Lohmann, and B. Wirnitzer. "Phase and Amplitude Recovery from Bispectra," Applied Optics: 3121-3129 (15 September 1984).

[2] Boll, Steven F. "Speech Enhancement in the 1980s: Noise Supression with Pattern Matching", in Advances in Speech Signal Processing. Eds. S. Furui and M.M. Sondhi. New York: Marcel Dekker, 1992.

[3] Ephraim, Yariv. "Statistical-Model-Based Speech Enhancement Systems," Proceedings of the IEEE, vol. 80, no. 10: 1526-1555 (October 1992).

[4] Fulchiero, Ralph and Andreas S. Spanias. "Speech Enhancement Using the Bispectrum," Proceedings of the International Conference an Acoustic, Speech and Signal Processing: IV-488 - IV-491 (1993).

[5] Kay, Stephen M. Modern Spectral Estimation - theory and application. New Jersey: PTR Prentice Hall Signal Processing Series, 1988.

[6] Lim, Jae S. and Alan V. Oppenheim. "All-Pole Modeling of Degraded Speech," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-26, no. 3: 197-210 (June 1978).

[7] Lohmann, Adolf W. and Bernhard Wirnitzer. "Triple Correlations," Proceedings of the IEEE, vol. 72 no. 7: 899-901 (July 1984).

[8] Makhoul, John. "Linear Prediction: A Tutorial Review," Proceedings of the IEEE, vol. 53 no. 4: 561-580 (April 1975).

[9] Matsuoka, Toshifumi and Tad J. Ulrych. "Phase Estimation Using the Bispectrum," Proceedings of the IEEE, vol. 72, no. 10: 1403-1411 (October 1984).

[10] Mendal, Jerry M. "Use of Higher-Order Statistics in Signal Processing and System Theory: An Update," SPIE Vol. 975 - Advanced Algorithms and Architectures for Signal Processing III: 126-144 (1988). {- not really used except to find further references}

[11] Mendal, Jerry M. "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretcial Results and Some Applications," Proceedings of the IEEE, vol. 79, no. 3: 278-305 (March 1991).

[12] Nikias, Chrysostomos L. "Higher-Order Spectral Analysis," Advances in Spectrum Analysis and Array Processing volume 1. Simon Haykin, editor. New Jersey: Prentice Hall, 1991.

[13] Nikias, Chrysostomos L and Mysore R. Raghuveer. "Bispectrum Estimation: A Digital Signal Processing Framework," Proceedings of the IEEE, vol. 75 no. 7: 869-891 (July 1987).

[14] Papoulis, Athanasios. Probability, Random Variables and Stochastic Processes. McGraw-Hill Series in Electrical Engineering, 1991.

[15] Rabiner, L.R. and R.W. Schafer. Digital Processing of Speech Signals. New Jersey: Prentice-Hall, Inc., 1978.

[16] Raghuveer, Mysore R. and L. Nikias Chrysostomos. "Bispectrum Estimation: A Parametric Approach," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol 33, no. 4: 1213-1229 (October 1985).

[17] Raghuveer, Mysore R. and L. Nikias Chrysostomos. "Bispectrum Estimation via AR Modeling," Signal Processing, vol. 10: 35-48 (1986).

[18] Parsons, Thomas W. Voice and Speech Processing. McGraw-Hill, Inc. (1987).

[19] Seetharaman, S. and M.E. Jernigan. "Speech Signal Reconstruction based on Higher Order Spectra," IEEE Transactions: 703-706 (1988).

[20] Sundarmoorthy, Gopal, et al. "Bispectral Reconstruction of Signals in Noise: Amoplitude Reconstruction Issues," IEEE Transactions of Acoustic Speech and Signal Processing, vol. 38 no. 7: 1297-1306 (July 1990).

[21] Wells, Benjamin B. "Voiced/Unvoiced Decision Based on the Bispectrum," Proceedings of the International Conference an Acoustic, Speech and Signal Processing: 1589-1592 (March 1985).

[22] Yannakoudakis, E.J. and P.J. Hutton. Speech Synthesis and Recognition Systems. Ellis Horwood Limited, 1987.

VITA

Major Deborah A. Douglass was born on 3 October 1963 in Moncton, New Brunswick. She graduated from University of Calgary in 1985 with a Bachelor of Science in Electrical Engineering. Upon graduation, Major Douglass received a commission in the Canadian Armed Forces as a Communications and Electrical Engineer. She began at a communications collection station at Canadian Forces Station Leitrum after which she moved through various other positions. In 1990, Major Douglass took a post as the Range Technical Officer at the Aerospace Engineering Test Establishment at Canadian Forces Base Cold Lake responsible for all ground-based test equipment. In May 1993, Major Douglass entered the School of Engineering, Air Force Institute of Technology at Wright-Patterson Air Force Base, Ohio to pursue a Master of Science (Space Operations).

Permanent Address: 1323 - 56th Avenue NW
Calgary, Alberta
Canada
T2K 5M2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1994	Masters Thesis

4. TITLE AND SUBTITLE

Voice Analysis using the Bispectrum

5. FUNDING NUMBERS

6. AUTHOR(S)

D.A. Douglass, Major
Canadian Armed FOrces

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GSO/ENG/94D-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

RL/IRAA
Rome Laboratory
32 Hanger Road
Griffiss AFB, NY 13441-4114

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The theory of the bispectrum has been studied, though very few practical applications have yet been considered in any depth. One application mentioned in the literature is the use of the bispectrum for voice signal processing. The aim of this thesis was to research the bispectrum towards the particular application of speech enhancement. The technique is based on the fact that the bispectrum is zero for a Gaussian white noise signal, and the bispectrum of two signals added together is the sum of the two signal bispectra. Theoretically, processing signals in the bispectra domain should increase the signal-to-noise ratio of the seech signal. The signal can then be reconstructed from the bispectrum.

Though the theory of the estimation techniques were proven, the applicability of the bispectrum to voice processing was questionable. Since any additive white noise is a random process, it will only be the expected value that is zero. With speech signals, the signal is considered stationary for only approximately 20 milliseconds. This does not allow a significant amont of the noise energy to be removed through the averaging process. Classical methods are just as effective.

14. SUBJECT TERMS

bispectrum, speech processing

15. NUMBER OF PAGES

70

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	DU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102